An Acoustic Charge Transport Imager for High Definition Television Applications

NASA Grant #NAGW-2753
Semi-Annual Report

W. D. Hunt and K. F. Brennan
School of Electrical and Computer Engineering
and
C. J. Summers
Georgia Tech Research Institute
Georgia Institute of Technology
Atlanta, Georgia 30332-0250

October, 1994

1.0 Program Objectives and Approach 1

2.0 Heterojunction Acoustic Charge Transport (HACT) Device 2
  2.1 New HACT Architectures and Established Region of Operation 2
    2.1.1 An Analytic Theory of Charge Transport in HACT Devices 3
    2.1.2 Experimental Technique for Transport Model Verification 16
    2.1.3 Development of the Injection Model 22
    2.1.4 Experimental Technique for Injection Model Verification. 36
    2.1.5 Summary of Results and Conclusions 43
  2.2 Piezoelectric Thin Films on GaAs-Based materials 44
    2.2.1 Background: 45
    2.2.2 ZnO film growth 46
    2.2.3 Characterization of acoustic properties 48
    2.2.4 Velocity surface 56
    2.2.5 Reflection properties of metallic gratings 66
    2.2.6 Summary of Thin Film Work 76
  2.3 Acoustoelectric Amplifiers 81
    2.3.1 Bilevel Model 81
    2.3.2 Segmented Amplifier 84
    2.3.3 ZnO Overlay 87
    2.3.4 Application to HACT Imager 88
    2.3.5 Summary 88
  2.4 Advances in ACT Device Modelling 91
    2.4.1 GaAs Heterojunction ACT Optimization 91
    2.4.2 ZnO on Silicon Acoustic Charge Transport 95
    2.4.3 Future Work on Device Simulation 95
  2.5 Materials Growth 98
    2.5.1 Substrate Heater Modification 98
    2.5.2 System Super Cleaning 99
    2.4.3 Future Work on Device Simulation 95
  2.5 Materials Growth 98
    2.5.1 Substrate Heater Modification 98
    2.5.2 System Super Cleaning 99
    2.5.3 Gallium and Aluminum Controller Replacement 99

3. Charge Transfer and Overflow Device 106
  3.1 Introduction 106
  3.2 Review of Charge Storage Devices 106
    3.2.1 General Concepts for Charge Storage Device 106
    3.2.2 Charge Storage Devices in GaAs 109
  3.3 Results 111
    3.3.1 Comparison Between Simulations and Experimental Results for CTD-5 111
    3.3.2 Device Characterization of CTD 115
    3.3.3 Fast Storage Lifetime Testing System for CTD 116

3.3.4 Summary and Future Work 117

4.0 Avalanche Photodiode Development 118
4.1 Theory of Impact Ionization and Modeling of Avalanche Photodiodes 118
4.2 Characterization of Avalanche Photodiodes 121

Bibliography 134

1.0 Program Objectives and Approach

The primary goal of this research is to develop a solid-state high definition television (HDTV) imager chip operating at a frame rate of about 170 frames/sec at 2 Megapixels per frame. This imager offers an order of magnitude improvement in speed over CCD designs and will allow for monolithic imagers operating from the IR to the UV.

The technical approach of the project focuses on the development of the three basic components of the imager and their integration. The imager chip can be divided into three distinct components: 1) image capture via an array of avalanche photodiodes (APDs), 2) charge collection, storage and overflow control via a charge transfer transistor device (CTD) and 3) charge readout via an array of acoustic charge transport (ACT) channels. The use of APDs allows for front end gain at low noise and low operating voltages while the ACT readout enables concomitant high speed and high charge transfer efficiency. Currently work is progressing towards the development of manufacturable designs for each of these component devices.

In addition to the development of each of the three distinct components, work towards their integration is also progressing. The component designs are considered not only to meet individual specifications but to provide overall system level performance suitable for HDTV operation upon integration. The ultimate manufacturability and reliability of the chip constrains the design as well.

The progress made during this period is described in detail in Sections 2-4.

2.0 Heterojunction Acoustic Charge Transport (HACT) Device

In this section of the report we describe the progress made in the development of the ACT component of the imager chip. In general, this includes:

1) The development and experimental investigation of new HACT architectures and the establishment of a Region of Operation curve, further, comprehensive.

2) The completion of a comprehensive study of the piezoelectric thin film ZnO on GaAs-based materials for HACT applications.

3) Initial studies into acoustoelectic amplifiers which are expected to drastically reduce the power budget for ACT devices.

4) Advances in ACT device modeling which allow us to simulate device performance in any of the competing ACT architectures.

2.1 New HACT Architectures and Established Region of Operation

This section presents a physically-based analytic model for charge injection and transport in HACT devices. Previous models have utilized numerical solutions of the coupled Poisson and current continuity equations [2.1.1-2.1.3]. This work, which was initially presented in [2.1.4], is the first physically-based analytic description of the charge injection process for ACT devices. Analytic models for charge transport in buried channel ACT devices were investigated by Hoskins, *et al.* [2.2.5]. This work extends this by refining some of the assumptions made by Hoskins, and by applying his approach to HACT devices. In the derivation of this model, bounds on different material parameters are developed that define the range under which acoustic charge transport may occur. This model is verified experimentally. Because the conditions for charge transport must exist before injection can occur, this is treated first. Expressions for the equilibrium and nonequilibrium carrier concentration are then developed. From this, an expression for the instantaneous channel current is then derived. The actual shape of the injection aperture has been the subject of some controversy [2.1.6], [2.1.7], [2.1.1], [2.1.8]. It is found and verified experimentally that the shape is approximately Gaussian. The form of the steady-state channel current has also been debated [2.1.6], [2.1.1], [2.1.8]. It is shown that the steady-state channel current has an exponential dependence on gate voltage.

The resulting models can be used to predict both the steady-state and transient characteristics of a HACT device of a particular geometry and material architecture. Being physically-based, generalizations about HACT device operation and trade-offs in device and material design can be made. Lastly, the model lends itself to a simple implementation on commercially available circuit analysis software.

2.1.1 An Analytic Theory of Charge Transport in HACT Devices

Hoskins and Hunsinger developed the first analytic theory of charge transport in ACT devices in [2.1.5]. This theory describes the transport of charge by the acoustically induced potential wave as a potential screening problem. In this section, this theory is extended and applied to heterostructure ACT devices. Different material constraints are identified, and their impact on the transport process is quantified. The model is also experimentally verified.

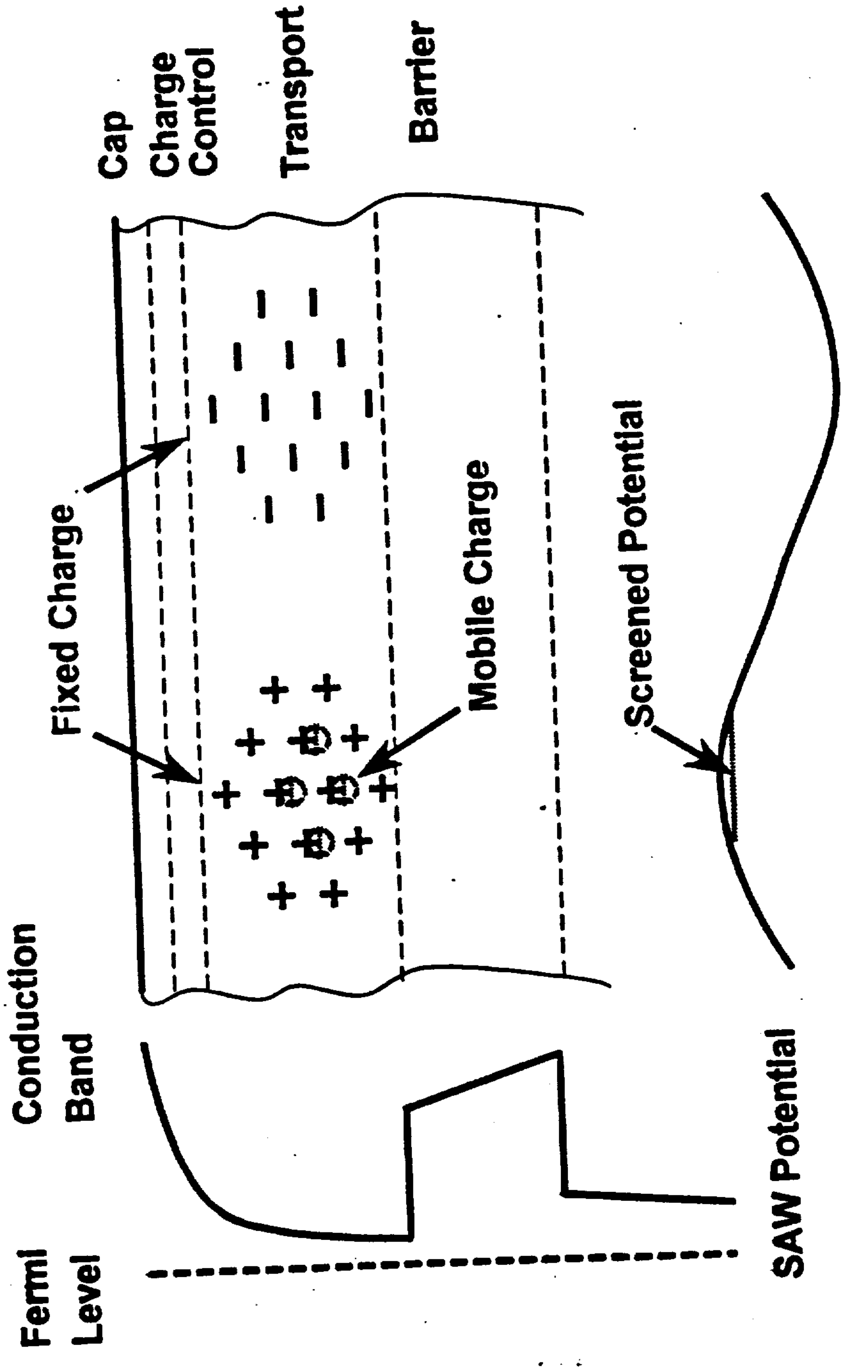
Cap
Charge Control
Transport
Barrier
Fixed Charge
Mobile Charge
Screened Potential
Conduction Band
Fermi Level
SAW Potential

Figure 2.1.1 Potential screening in HACT devices.

2.1.1.1 The HACT Potential Screening Problem.

Acoustic charge transport is essentially a process whereby mobile electrons screen disturbances in the fixed charge concentration created by the motion of the SAW wave in the piezoelectric semiconductor [2.1.6]. Generally, this process is solved under the assumption of a stationary potential well, since the period of the SAW is much greater than the dielectric relaxation time of the medium [2.1.5]. The statement of the problem is then as follows: *Given an acoustically- induced charge disturbance which is assumed stationary and superimposed on a background of fixed charge, determine the equilibrium free carrier concentration which screens this charge and the distance over which it is distributed.*

This problem is illustrated in Figure 2.1.1, which also shows the effects of the various layers in a typical HACT material structure. The primary purpose of the heterostructure is to confine mobile charge in the vertical direction [2.1.7]. A *cap layer* is used to control the effects of surface states on the conduction band in this vertical direction. Beneath this, is the *charge control layer*, which acts to form the upper potential barrier. Two alternatives have been proposed for this layer. Tanski, *et al.* proposed an $Al_{.3}Ga_{.7}As$ layer which creates a potential discontinuity of approximately 0.25 eV in the conduction band [2.1.7]. Grudkowski, *et al.* later proposed a p-doped GaAs layer to create a barrier from the built-in potential of the depletion region [2.1.9]. Smith, *et al.* have also investigated this structure [2.1.10]. Ideally, the *transport layer* is the region where most of the mobile charge is confined. Beneath this, the *barrier layer* defines the lower region of charge confinement. A *buffer layer* is sometimes employed to minimize growth defects in the upper layers.

The potential to be screened is due to the superposition of the background donors and the polarization charge induced by the acoustic wave. The amplitude of the surface potential due to the acoustic wave, ϕ_{s0}, is related to the voltage, V_{tr}, applied to a transducer by

$$\phi_{s0} = 2\eta\,\xi\,\kappa^2 N V_{tr} \tag{2.1.1}$$

where η is a parameter dependent on the metalization ratio ($\approx$0.8 for equal gap and width size), ζ is the electromechanical conversion efficiency (~0.7 typical), κ^2 is the piezoelectric coupling constant (=0.0007 for GaAs) and N is the number of finger pairs in the IDT [2.1.11]. The factor of two arises because the use of a reflector is assumed, which doubles the amplitude of the SAW. A common approximation assumes that the acoustic wave emanating from the IDT is an infinite plane wave in the transverse direction [2.1.12]. As a result of this approximation, the variation of the potential in the direction perpendicular to the transport direction is assumed to be zero. A similar, but somewhat cruder approximation is made with regard to the potential variation into the substrate. The amplitude of the potential induced by a SAW on an electrically open surface is shown to fall off approximately exponentially with the acoustic wavelength into the substrate [2.1.11]. However, because the transport channel layer is usually a small fraction of an acoustic wavelength, the variation in potential in this direction can be neglected. Furthermore, the depth of the channel is generally small enough that the potential at the transport depth is essentially the same as at the surface, as given by equation (2.1.1). The SAW can therefore be expressed by an effective charge concentration, N_{s0}, related to the SAW potential by the one-dimensional Poisson's equation

$$\frac{d^2}{dx^2}\left[\phi_s \cos\left(\frac{2\pi x}{\lambda_a}\right)\right] = \frac{qN_{s0}}{\epsilon_s}\cos\left(\frac{2\pi x}{\lambda_a}\right) \tag{2.1.2}$$

where λ_a is the acoustic wavelength, and $N_{s0} = 4\pi^2\phi_s\varepsilon_s/q\lambda_a^2$ is the peak polarization charge concentration, with q being the charge on an electron, and ε_s the permittivity of the material.

The equilibrium electron concentration can be accurately obtained by self consistently solving the Poisson equation coupled to the Boltzmann distribution using numerical methods [2.1.13]. This has been done by a number of authors [2.1.1-2.1.3, 2.1.10]. However, insight into the physics of the device can also be gained by analyzing the asymptotic solutions of the nonlinear ordinary differential equation derived from directly substituting the Maxwell-Boltzmann distribution in Poisson's equation [2.1.14]. The free electron concentration, n, can be expressed as a function of the conduction band potential $\psi(x) = E_c(x)/q$, where E_c is the conduction band energy relative to the degeneracy level,

$$n(x) = N_c e^{-\frac{q}{kT}\psi(x)} \tag{2.1.3}$$

where k is Boltzmann's constant, T is the absolute temperature, and N_C is the density of states in the conduction band. Assuming a transport layer doping, N_{Dch}, and neglecting any bandgap narrowing effects, equation (2.1.3) can be substituted in Poisson's equation, to yield

$$\frac{d^2\psi}{dx^2} = \frac{\rho}{\epsilon_s} = \frac{q}{\epsilon_s}\left[N_{Dch} + N_{s0}\cos\left(\frac{2\pi x}{\lambda_a}\right) - N_c e^{-\frac{q\psi}{kT}}\right] \tag{2.1.4}$$

where ρ is the charge density in the region of interest. equation (2.1.4) is a nonlinear ordinary differential equation that is not easily solved by analytic methods. To quantify the potential, and hence the concentration throughout the transport channel, the domain over which equation (2.1.4) is solved is broken up into two distinct regions. Over regions where there is no gradient in potential (i.e. the "flat band" condition), equation (2.1.4) reduces to an algebraic expression, and the region is said to be *quasi-neutral*. This implies perfect screening of the SAW potential and donor ions by the free electrons. Hence,

$$n(x) = N_{Dch} + N_{s0}\cos\left(\frac{2\pi x}{\lambda_a}\right) \ . \tag{2.1.5}$$

This implies that the potential within the charge packet is

$$\psi(x) = \frac{kT}{q}\ln\left[\frac{N_c}{N_{Dch} + N_{s0}\cos\left(\frac{2\pi x}{\lambda_a}\right)}\right] \qquad 0 \le x < x_0 \ . \tag{2.1.6}$$

where x_0 defines the charge packet boundary. Hoskins, *et al.* arrived at the above result, but assumed that the potential outside the packet boundary was equal to the SAW potential [2.1.5]. This is true only for very low doped channels. Schmuckler later developed an empirical model for the problem of barrier lowering by the packet charge [2.1.15]. This work extends the previous work of others by developing a physically-based expression for the potential outside the charge packet by assuming that the *depletion approximation* holds whereby n is assumed negligible compared to the doping and polarization charge. In this case, the linear Poisson equation

$$\frac{d^2\psi}{dx^2} = \frac{q}{\epsilon_s}\left[N_{Dch} + N_{s0}\cos\left(\frac{2\pi x}{\lambda_a}\right)\right] \tag{2.1.7}$$

can be easily solved to arrive at an analytic expression for the potential. This can be accomplished by direct integration, yielding

$$\psi(x) = \frac{1}{2}\frac{q}{\epsilon_s}N_{Dch}(x - x_0)^2 + \phi_{s0}\cos\left(\frac{2\pi x}{\lambda_a}\right) + A(x - x_0) + B \qquad x_0 \leq x < \lambda_a/2 \tag{2.1.8}$$

where A and B are constants of integration. These are evaluated by requiring continuity of the potential function and its derivative at x_0. In doing so,

$$A = \left[\frac{kT}{q}\cdot\frac{N_{s0}}{N_{Dch} + N_{s0}\cos\left(\frac{2\pi x_0}{\lambda_a}\right)} + \phi_{s0}\right]\frac{2\pi}{\lambda_a}\sin\left(\frac{2\pi x_0}{\lambda_a}\right)$$

$$B = \frac{kT}{q}\ln\left[\frac{N_c}{N_{Dch} + N_{s0}\cos\left(\frac{2\pi x_0}{\lambda_a}\right)}\right] - \phi_{s0}\cos\left(\frac{2\pi x_0}{\lambda_a}\right) \tag{2.1.9}$$

The only parameter left to determine to complete the approximate analytic solution to the potential is the length of the packet boundary x_0. Hoskins, *et al.* obtained a simple form of this screening length by assuming that the concentration drops to zero outside the region where the sum of the SAW potential and the ionized dopant potential goes negative [2.1.5]. Hoskins' model therefore assumes that charge can be screened, and hence transported in an undoped channel. In HACT devices, it has been experimentally determined that this is not the case. This will be discussed in Section 2.1.3. A physical basis for this can be derived from Hutson and White's theory of acoustoelectric attenuators and amplifiers: *Charge screening in HACT devices occurs only in regions where the spatial variation of the SAW potential is small on a Debye length scale within the channel* [2.1.16], [2.1.17]. Thus, when the Debye length becomes an appreciable fraction of the acoustic wavelength, the polarization charge is not effectively screened by the free electrons in the channel. This forms a basis to determine x_0. Debye length is a function of free carrier concentration and is defined by the length over which a small ($<<kT/q$) perturbation from the flat band potential decays to $1/e$ of the initial value [2.1.14]. This results from the Taylor expansion of the exponential term in equation (2.1.3), and is not really appropriate for a large potential deviation. Considering this, a function, $L_D(x)$, termed the "quasi-Debye length," can be developed by substituting the packet concentration for the carrier concentration in the normal Debye length expression:

$$L_D(x) = \sqrt{\frac{\epsilon_s kT}{q^2\left[N_{Dch} + N_{s0}\cos\left(\frac{2\pi x}{\lambda_a}\right)\right]}} \ . \tag{2.1.10}$$

The screening length x_0 is defined by the condition $L_D(x_0) = L_{D0}$, where L_{D0} is an empirical parameter expected to be a small fraction of the acoustic wavelength λ_a. In doing this, the screening length is given by

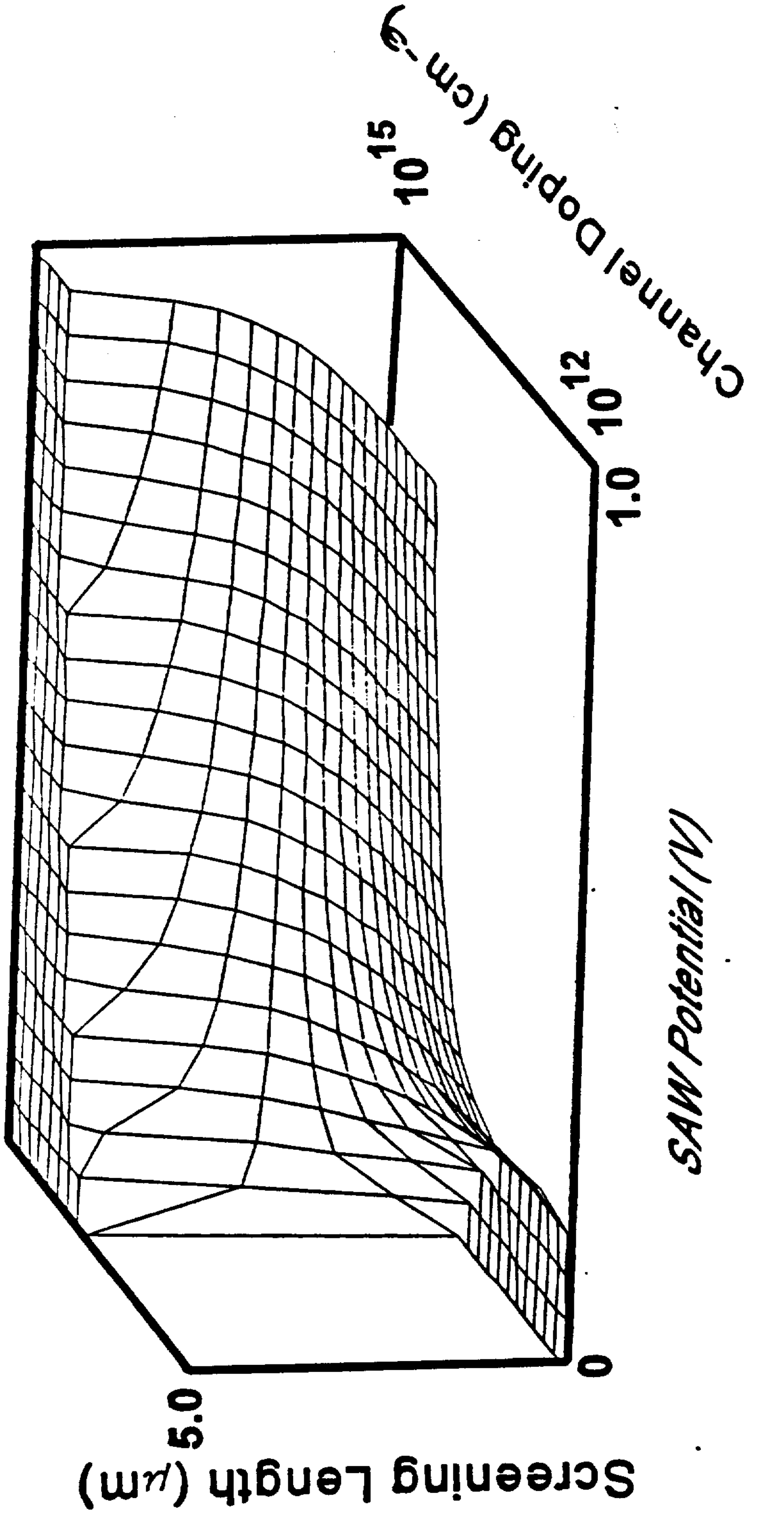

Figure 2.1.2 Predicted charge screehing length (packet length) as a function of SAW potential and channel doping.

$$x_0 = Re\left\{\frac{\lambda_a}{2\pi}\cos^{-1}\left[\frac{\epsilon_s kT}{q^2 L_{D0}^2 N_{s0}} - \frac{N_{Dch}}{N_{s0}}\right]\right\} \qquad (2.1.11)$$

Figure 2.1.2 shows the predicted screening length (i.e. the length of the charge packet) versus SAW potential and channel doping for the material design presented in [2.1.10] . It is seen that there are local regions where the screening length is zero, indicating the charge transport will not occur. For this material design, the SAW potential threshold for charge transport to occur is on the order of 0.25 V in low doped channels.

2.1.1.2 Material Constraints and Operating Conditions for Efficient Transport.

Eqs. (2.6, 2.8, 2.9 and 2.11) form a complete analytic model for charge transport in HACT devices. The determination of L_{D0} must be done empirically. The accuracy of this model is verified by comparing the predicted potential with that calculated by numerically solving equation (2.1.3). This is shown in Figure 2.1.3. The Runge-Kutta method was used to solve equation (2.1.3) over one half of an acoustic wavelength [2.1.18]. It is seen that the analytic theory is close to that predicted by the numerical method. A value of $L_{D0} = \lambda_a/18$ was found to give the closest fit for lower doped channels, while $L_{D0} = \lambda_a/5$ gave a better fit for higher doped channels. It is also apparent that increasing the channel doping has two effects: *increasing the well charge capacity, and decreasing the potential barrier between wells.* Obviously, when $N_{Dch} = N_{s0}$, the well is essentially full. An approximate expression for the full well charge capacity is easily derived by integrating equation (2.1.5) over the effective screening length given in equation (2.1.11):

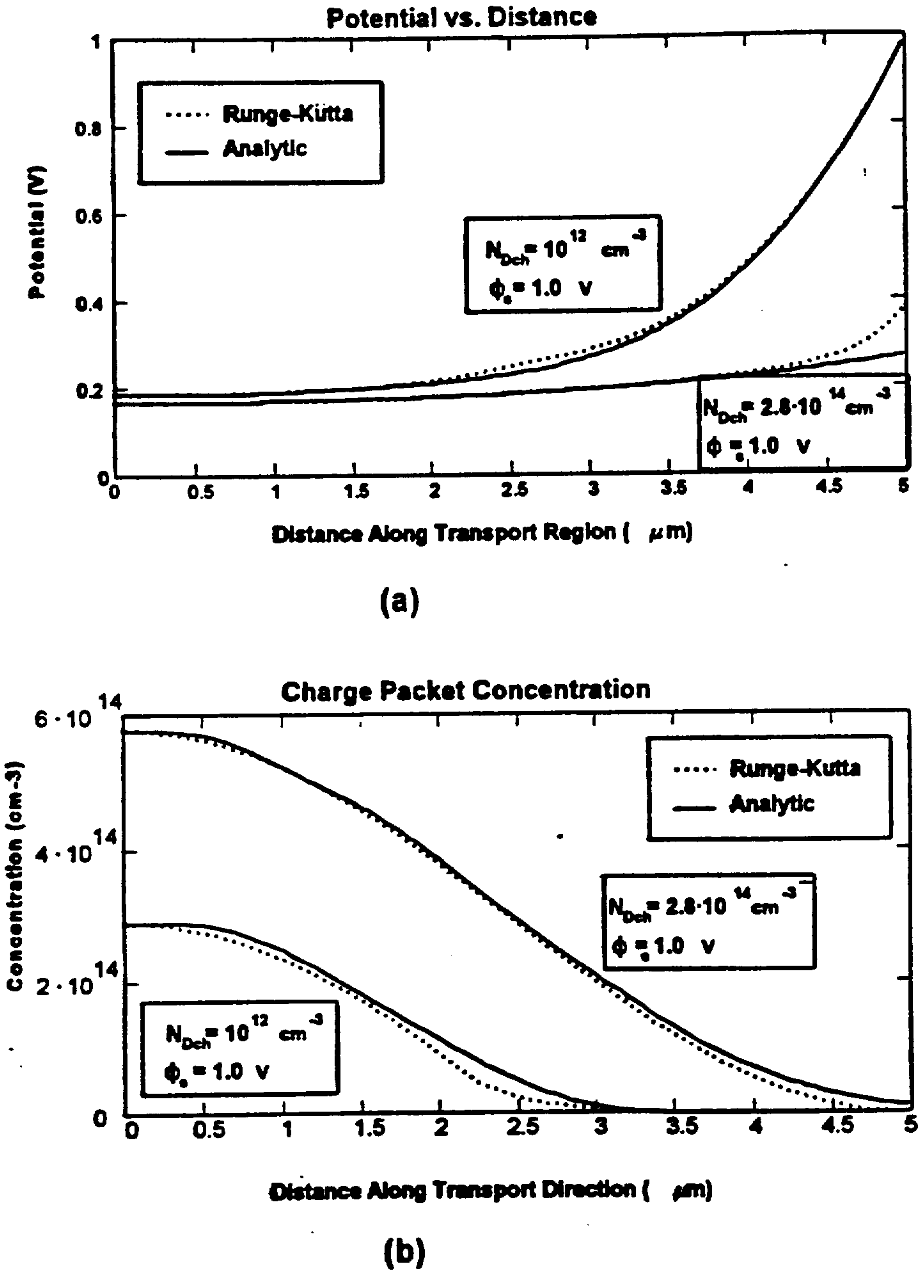

Figure 2.1.3 (a) Calculated potential over one half wavelength for two different channel dopings. (b) Free electron concentration calculated from potential.

$$N_F = 2Wd_{ch}\left[N_{Dch}x_0 + \frac{2\pi\epsilon_s}{q\lambda_a}\phi_{s0}\sin\left(\frac{2\pi x_0}{\lambda_a}\right)\right] \tag{2.1.12}$$

where the charge outside the screening length is neglected, and the concentration is assumed constant over the channel layer. The later assumption is only valid over the region where the channel is quasi-neutral.

Hoskins, *et al.* pointed out that a trade-off must be made between the well capacity and the diffusion induced transfer inefficiency [2.1.5]. The analytic model just developed provides a convenient means to predict the barrier height, and hence the CTE. Taking the barrier height as the difference between the potential minimum at x=0, and maximum at $x=\lambda_a/2$, the well barrier height ϕ_w is given by

$$\phi_w = \frac{qN_{Dch}}{2\epsilon_s}\left(\frac{\lambda_a}{2} - x_0\right)^2 - \phi_{s0} + A\left(\frac{\lambda_a}{2} - x_0\right) + B - \frac{kT}{q}\ln\left(\frac{N_c}{N_{Dch} + N_{s0}}\right) \tag{2.1.13}$$

A plot of the barrier height versus channel doping is shown in Figure 2.1.4 for several values of SAW potential. Hoskins used thermionic emission theory to predict the charge loss per clock cycle [2.1.5]. The results of his work indicate that barrier height of less than about 5*kT* may result in a significant loss of charge in typical device architectures. Figure 2.1.4 also points out several areas in which this theory improves upon models suggested in [2.1.5] and [2.1.6]. The maximum barrier height is somewhat lower than that predicted by these authors because of the inclusion of the effects of finite Debye length. For the same reason, the optimum channel doping is not equal to the polarization charge amplitude as they suggested. The optimum channel doping

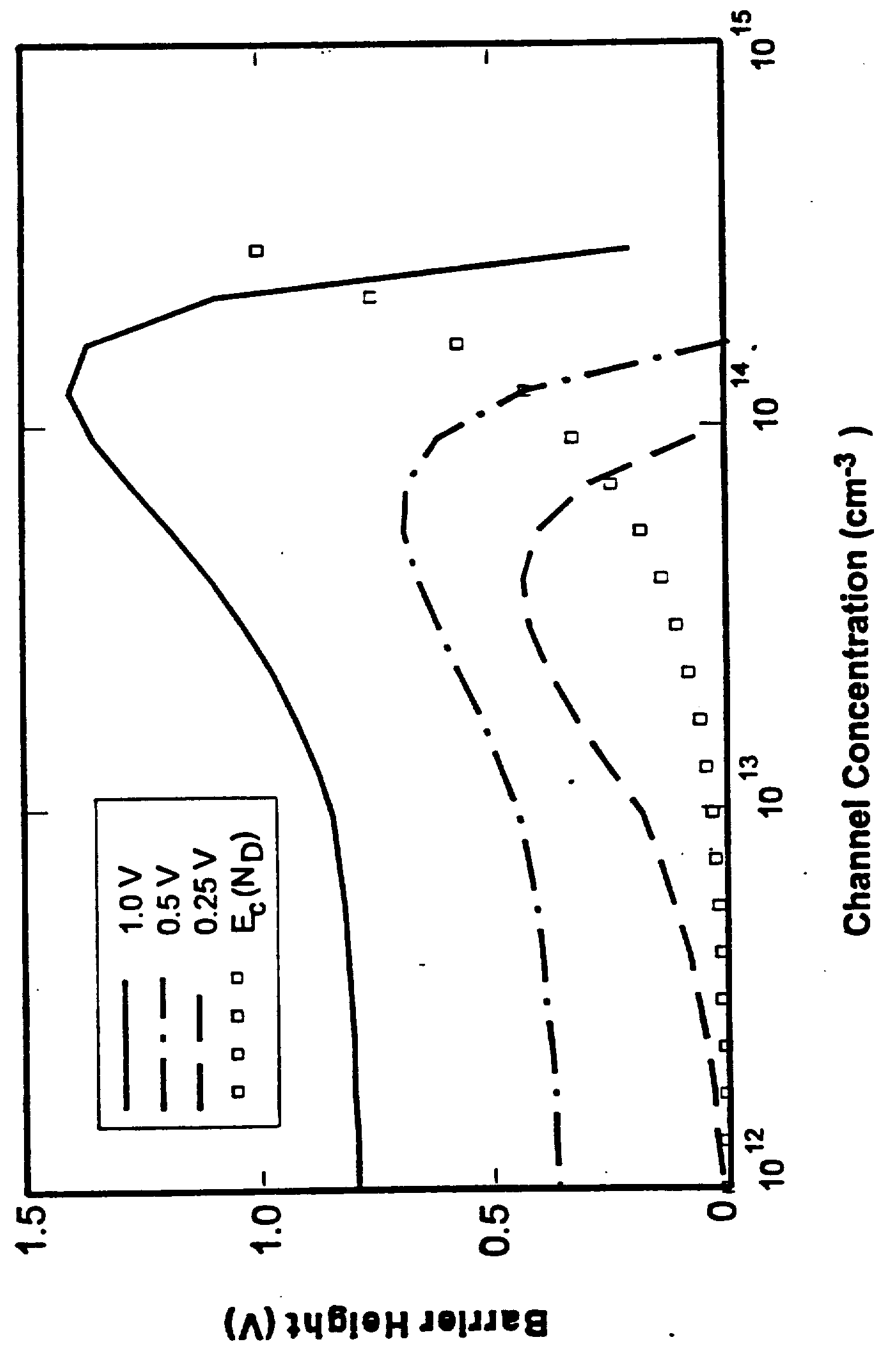

Figure 2.1.4 Height of potential barrier between adjacent packets for several channel dopings. Polarization charge amplitude for a given SAW potential is indicated by the squares.

would lie on a line passing through the peaks of the curves shown in Figure 2.1.4. This is seen to be somewhat less than the polarization charge amplitude.

2.1.2 Experimental Technique for Transport Model Verification

The model developed in the last section was verified to some extent by comparing the predictions of the analytic expression to numerically derived results. However, to truly assess the validity of the model, the predicted values of several parameters must be compared to those obtained by measurements on actual devices. Basically, the transport model predicts the conditions under which acoustic charge transport will occur in HACT devices. It will be shown in this section that devices operating outside a range of SAW potential and equilibrium channel concentration are not measurably functional, whereas those that fall within this range show measurable ACT effects. Furthermore, equation (2.1.12) allows the prediction of charge capacity. This will also be verified experimentally.

2.1.2.1 Test Setups and Procedures.

In order to verify that charge is being transported by the acoustic potential, two things must occur: 1) *Channel current must be present only when the SAW potential is applied*, and 2) *A transient signal imposed on the input must be delayed by the time it takes the acoustic wave to traverse the channel.* Figure 2.1.5 shows the setup used to measure both DC channel current and frequency response. While DC channel current is relatively easy to verify, the signal delay is sometimes not. Because of the low levels of the measured signal compared to spurious feedthrough signals, the expected frequency response due to the NDS array is badly distorted. However, it is possible to distinguish the acoustically delayed signal from the feedthrough signal

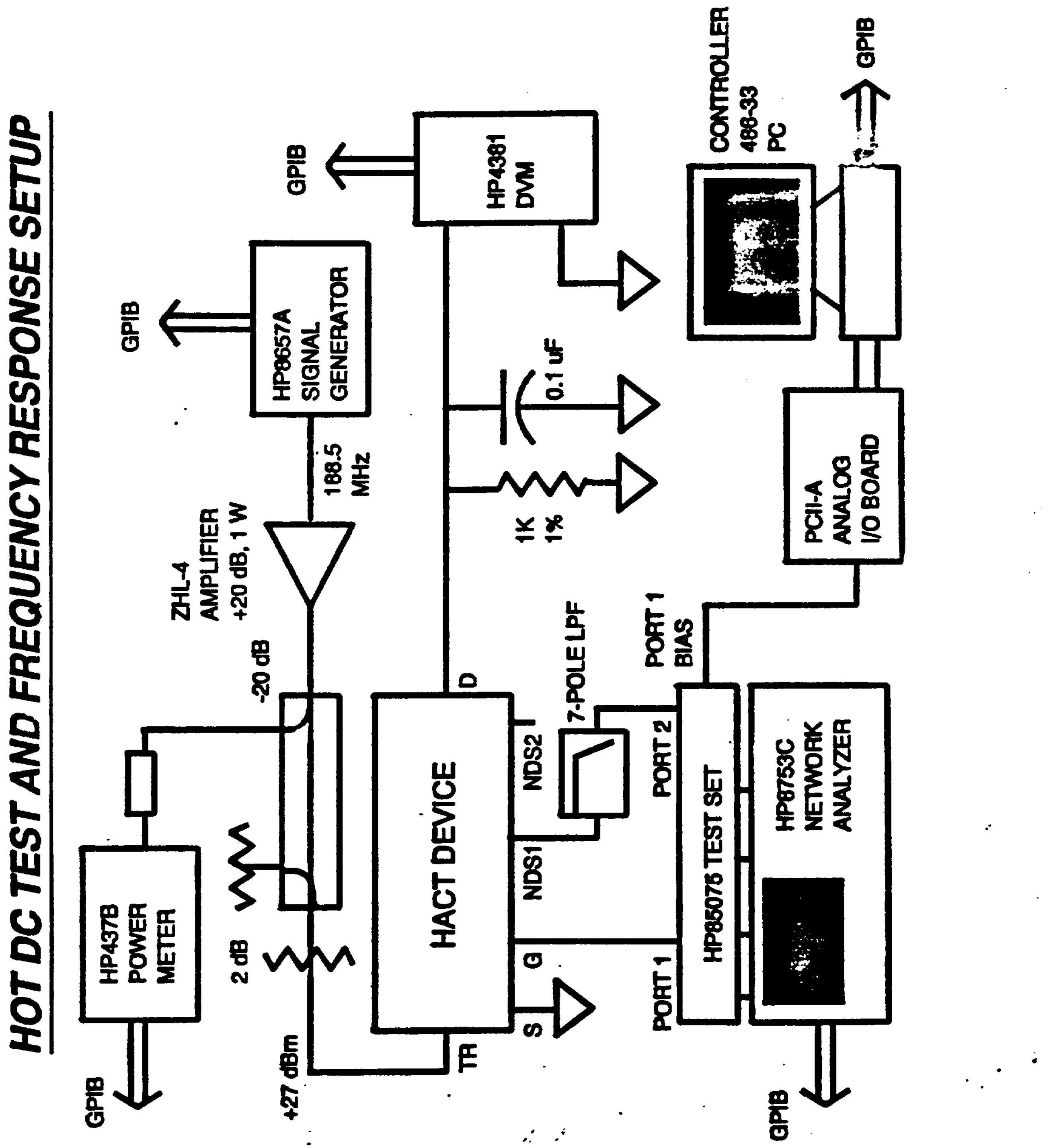
HOT DC TEST AND FREQUENCY RESPONSE SETUP
GPIB
HP437B
POWER
METER
2 dB
+27 dBm
-20 dB
ZHL-4
AMPLIFIER
+20 dB, 1 W
188.5
MHz
GPIB
HP8657A
SIGNAL
GENERATOR
GPIB
HP4381
DVM
TR
S
G
NDS1
NDS2
D
HACT DEVICE
1K
1%
0.1 uF
7-POLE LPF
PORT 1
PORT 2
PORT 1
BIAS
HP85075 TEST SET
HP8753C
NETWORK
ANALYZER
GPIB
PCII-A
ANALOG
I/O BOARD
CONTROLLER
486-33
PC
GPIB

Figure 2.1.5 Test setup used to measure DC channel current and frequency response.

by performing a Fourier transform on the measured frequency response. This is shown in Figure 2.1.6. It is seen that the feedthrough signal is of much greater amplitude than the acoustically delayed signal. However, if the output taps are spaced sufficiently far from the input gate, the feedthrough signal decays to the point where the acoustically delayed signal is measurable. Hence the ACT operation is verified.

2.1.2.2 Measurements and Results.

Figure 2.1.7 shows a summary of the performance of HACT devices fabricated at Georgia Tech and United Technologies versus the equilibrium channel concentration and SAW potential. This plot shows three different constrains on the channel concentration and SAW potential. Two of these, the Polarization Charge Limit and the Transport Limit are discussed in the section, while the Injection Barrier Limit is discussed in Section 2.1.3. A lower limit on charge transport is defined by the screening length, as given in equation (2.1.11), reducing to zero. Since this is dependent on frequency, the three different curves represent the three different operating frequencies of the different devices measured. Filled markers indicate device wafer lots in which charge transport was verified, and empty markers designate nonfunctional devices. It is seen that this limit is only relevant at low SAW potentials ($< 8kT/q$), and only for low channel dopings. It is seen that the condition is automatically met if the conditions are met for Schottky injection, which is discussed in the next section. Thus it was not possible to verify the lower boundary of HACT operation using the injection structures of the test devices.

Verification of the full-well charge capacity given by equation (2.1.12) was done by measuring the maximum channel current that the SAW can carry. Figure 2.1.8 shows the channel current as a function of gate voltage for United Technologies HACT B-3. This will be discussed

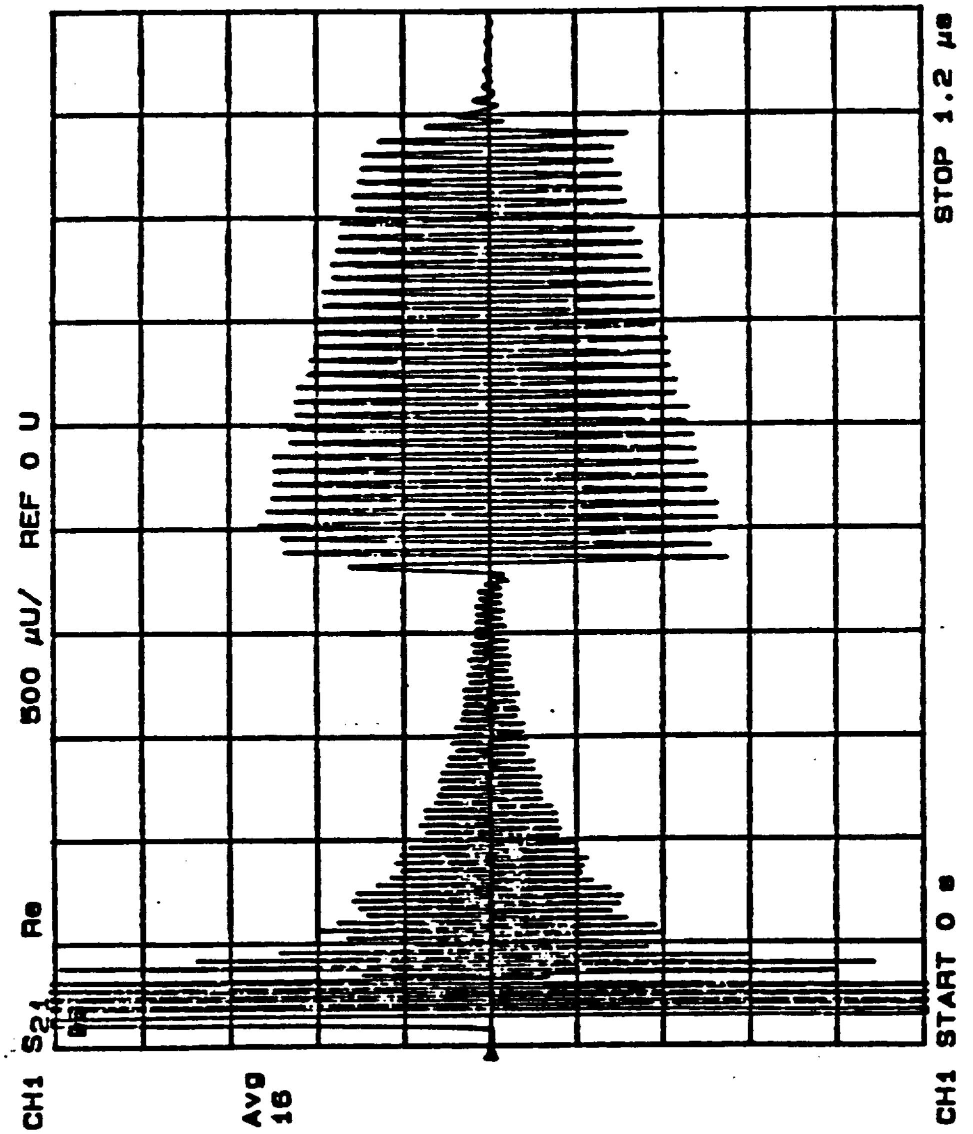

Figure 2.1.6 Output signal of device D02 of HACT lot 1854. Time domain signal is calculated by a Fourier transform of measured frequency response. Signal delay of 600 ns confirms ACT operation

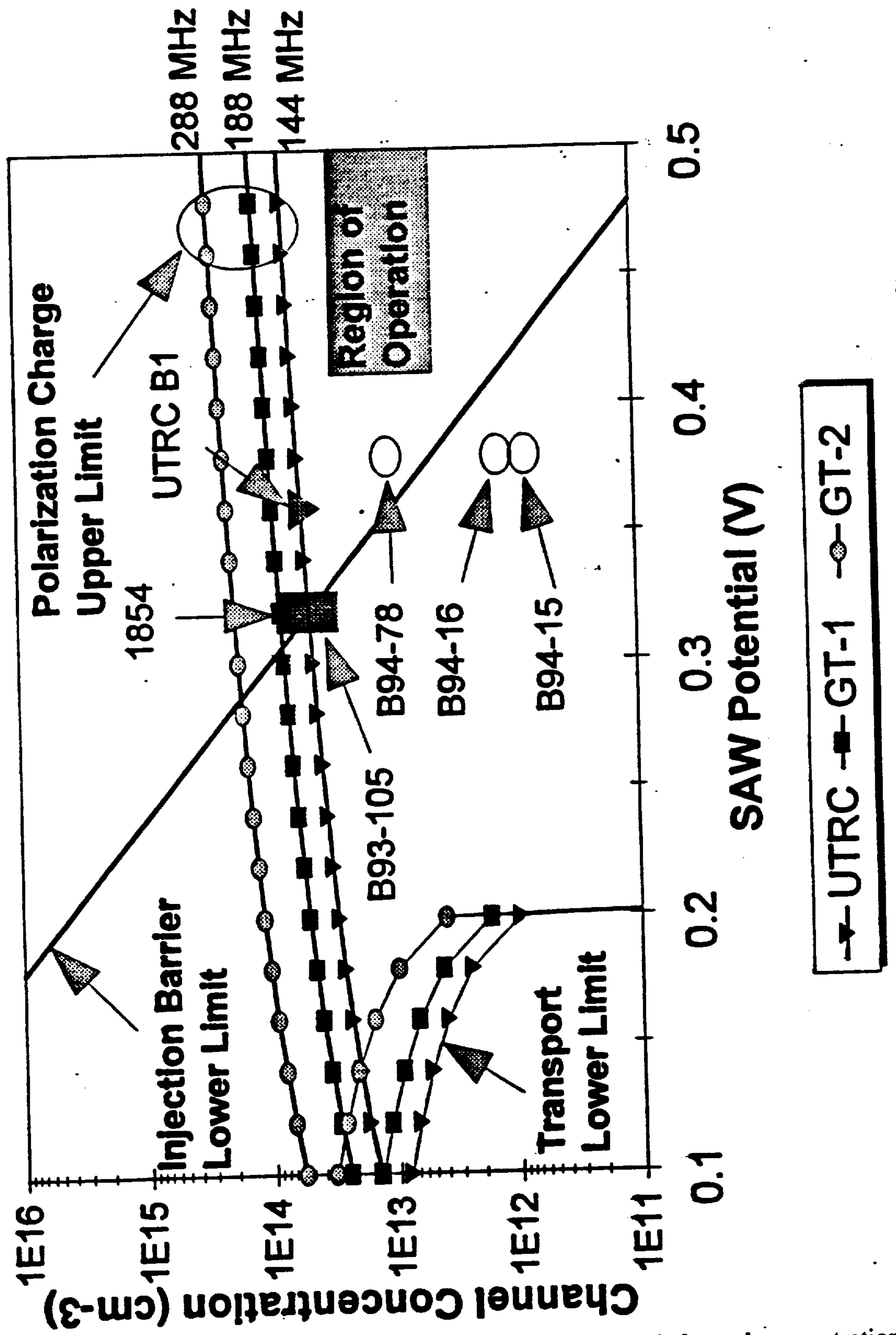

Figure 2.1.7 Functionality of HACT devices versus SAW potential and channel concentrations. Filled markers indicate that charge transport was verified, empty markers denote non-functional devices

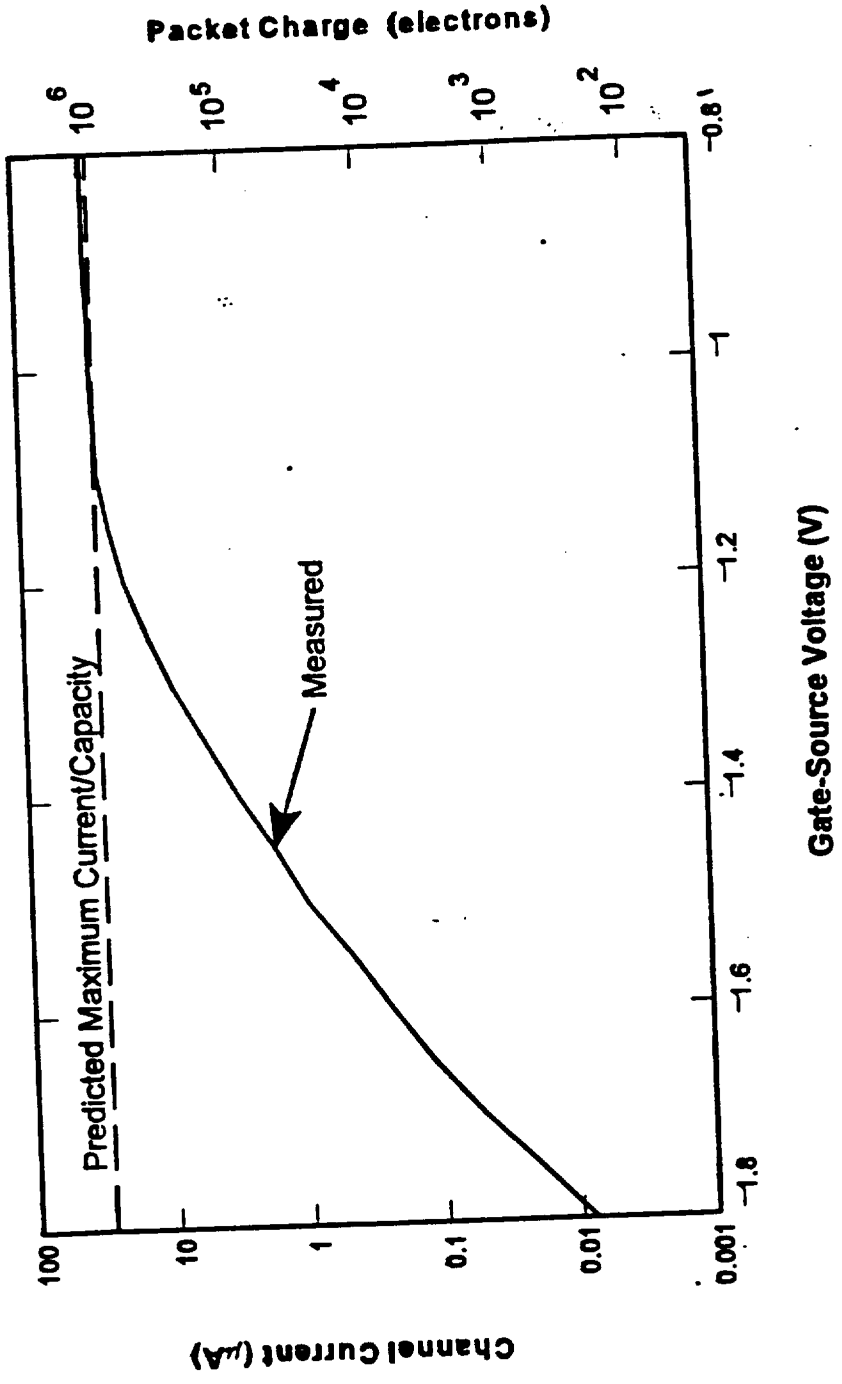

Figure 2.1.8 Measured average channel current on UTRC HACT B-1 compared to maximum, equivalent electron packet size is also shown.

in greater detail in the next section. It is seen that the channel current abruptly stops increasing as a function of gate voltage, indicating that the potential wells are at their peak capacity for the operating SAW potential. Since the average channel current is simply the product of the packet charge and the SAW frequency, the right axis is plotted in units of electrons per packet. It is seen that the maximum capacity is close to that predicted by equation (2.1.12).

2.1.3 Development of the Injection Model

As illustrated in Figure 2.1.9, the input structure superficially resembles a HEMT [2.1.19]. However, differences between the two devices are numerous. Under normal operation all charge is moved by the SAW potential travelling at the acoustic velocity, which is two orders of magnitude slower than the saturation velocity in a HEMT. The epitaxial layers also differ. In a HEMT, the AlGaAs charge control layer in heavily doped *n*-type. As described by the Anderson model, the differing electron affinities of the narrow bandgap GaAs and the wider gap AlGaAs create a conduction band discontinuity at the heterojunction interface [2.1.20]. In equilibrium the doped AlGaAs layer provides electrons which collect in the potential well formed by this discontinuity. This creates a highly concentrated 2-DEG channel region. In contrast, the charge control layer in a HACT acts only to form a space-charge region which broadly confines the carriers in the vertical direction the transport channel. Also, as discussed in the last section, the equilibrium channel concentration is kept quite low ($\sim 10^{14}$ cm^{-3}) so as not to exceed the polarization charge limit. Moreover, the cap layer of a HEMT structure is heavily doped *n*-type to aid in the formation of channel ohmic contacts. This is not possible in the HACT structure, because the shorting of the acoustic potential at the surface will cause a significant decrease in

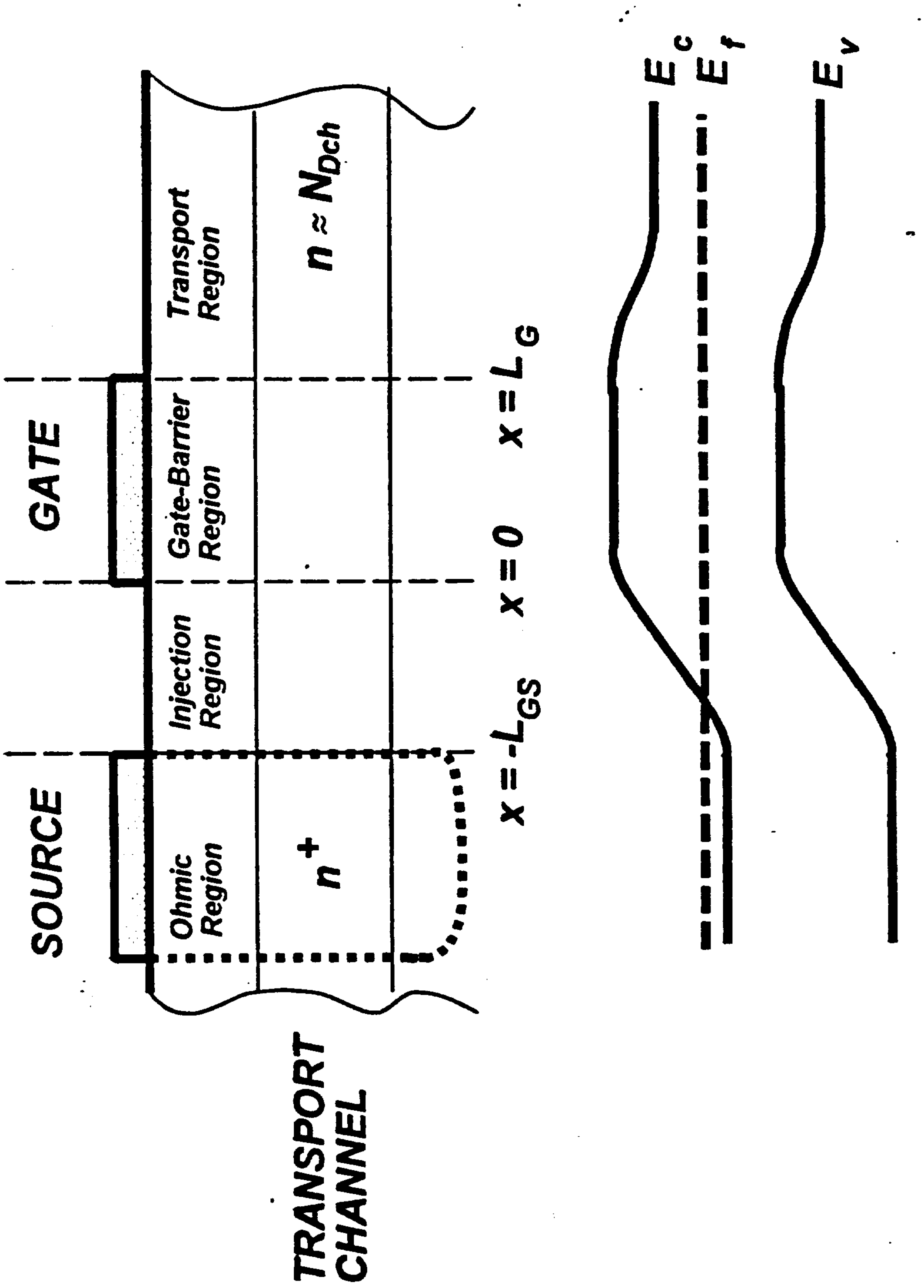
SOURCE
GATE
Ohmic Region
Injection Region
Gate-Barrier Region
Transport Region
n^+
$n \approx N_{Dch}$
$x = -L_{GS}$
$x = 0$
$x = L_G$
E_c
E_f
E_v
TRANSPORT CHANNEL

Figure 2.1.9 Schematic of the HACT injection structure and associated equilibrium band diagram.

the SAW potential at the transport depth [2.1.11]. Thus it turns out that most MESFET and HEMT equivalent circuit models are not really applicable to HACT devices. However, the subthreshold region of MESFET and HEMT operation, not generally considered in the development of equivalent circuit models, does resemble the charge injection into the HACT channel. When the channel region is completely pinched-off, the MESFET or HEMT is said to be in the *subthreshold* region [2.1.14], [2.1.21], [2.1.22]. In a conventional MESFET or HEMT operating in this regime, charge present in the n^+ ohmic region diffuses over the potential barrier set up by the gate voltage. It might also be noted that injection in HACT devices is similar to charge injection in heterostructure insulated gate FET's (HIGFET's) [2.1.23].

The theory of HACT charge injection is presented in the following subsections. The development this theory parallels the theory of subthreshold injection in more conventional semiconductor devices. The inclusion of the SAW potential does complicate the problem somewhat, so the derivation is broken up into several steps, the first of which is to determine the equilibrium mobile electron concentration throughout the injection region. Although numerical methods are available to do this accurately, some assumptions can be made that allow the formulation of an approximate analytical solution. The next step is to evaluate the nonequilibrium concentrations in the injection region. Again, simplifying assumptions are made to make the problem tractable using analytic methods. Once the mobile carrier concentrations are known, the final step is to incorporate these into the drift-diffusion equation in order to arrive at an analytical expression for the instantaneous channel current. An additional step is performed to obtain a closed-form expression for the steady-state channel current as a function of gate potential.

2.1.3.1 Determination of Equilibrium Carrier Concentration.

Referring to Figure 2.1.6, the injection region is divided into four parts. In the first region, the ohmic contact ($x<L_{GS}$) contains a high concentration of mobile electrons due to the diffusion of dopants during the alloy process. The determination of the actual free carrier concentration in this region is quite difficult for several reasons. The first of which is the degenerate nature of the doping in this region $N_{DS}(x)$, which precludes the use of Boltzmann statistics. Secondly, widely differing reported experimental results imply that the alloy process itself is not well understood [2.1.24]. It is often assumed that the concentration is on the order of 10^{19} cm^{-3}, and that the ohmic drop during the current injection is negligible [2.1.25]. Within the ohmic region, the conduction band is considered flat, and quasi-neutrality is assumed. Hence, $n \approx N_{DS}$, since the doping concentration is many orders of magnitude higher than the polarization charge. The potential is related to the carrier concentration by

$$n(x) = N_c F_{1/2}\left[-\frac{q}{kT}\psi(x)\right] \tag{2.1.14}$$

where $F_{1/2}$ is the Fermi integral of order 1/2 [2.1.14].

The second region, between the source and gate contacts ($-L_{GS} < x < 0$), will be referred to as the "injection region" throughout this section, as it is the concentration in this region that most greatly affects the charge injection operation. Here the concentration can be found from the solution of Poisson's equation. Since the potential in this region is reacting to a large change in doping, neither the quasi-neutrality nor depletion approximations can be assumed [2.1.14]. The equation to be solved is then

$$\frac{d^2\psi}{dx^2} = \frac{q}{\epsilon_s}\left[N_D(x) - N_c e^{-\frac{q}{kT}\psi}\right] \tag{2.1.15}$$

where $N_D(x) = N_{DS}$ for $x < -L_{GS}$ and $N_D(x) = N_{Dch}$ for $-L_{GS} < x < 0$. What is often done for small potential perturbations ($\Delta\psi << kT/q$) is to expand the exponential dependence of the carrier concentration on the potential in a Taylor series, resulting in a linear ordinary differential equation, and the well-known definition of Debye length [2.1.19]. This result can be applied to the more general problem by recognizing that a larger change will result in a similar exponential change in the potential. The form of this dependence can be deduced by empirically altering the solution to the linear problem as follows:

$$\psi(x) = \begin{cases} \psi_{0S}\left(1 - \alpha_1 e^{-\gamma_1 \frac{x - L_{GS}}{\lambda_{DS}}}\right) & x < -L_{GS} \\ \psi_{0ch}\left(1 - \alpha_2 e^{-\gamma_2 \frac{x - L_{GS}}{\lambda_{Dch}}}\right) & -L_{GS} \le x < 0 \end{cases} \tag{2.1.16}$$

where

$$\begin{aligned} \psi_{0S} &= \frac{kT}{q}\ln\left(\frac{N_c}{N_{DS}}\right), & \lambda_{DS} &= \sqrt{\frac{\epsilon_s kT}{q^2 N_{DS}}} \\ \psi_{0ch} &= \frac{kT}{q}\ln\left(\frac{N_c}{N_{Dch}}\right), & \lambda_{Dch} &= \sqrt{\frac{\epsilon_s kT}{q^2 N_{Dch}}} \end{aligned} \tag{2.1.17}$$

Requiring continuity of the potential and its derivative results in expressions for the parameters α_1 and α_2:

$$\alpha_2 = \frac{1 - \psi_{0ch}/\psi_{0S}}{1 + \dfrac{\lambda_{DS}\gamma_2}{\lambda_{Dch}\gamma_1}} \quad , \quad \alpha_1 = -\alpha_2 \frac{\psi_{0S}\lambda_{DS}\gamma_2}{\psi_{0ch}\lambda_{Dch}\gamma_1} \tag{2.1.18}$$

where γ_1 and γ_2 are determined empirically. One might expect that the potential will drop off somewhat less than the Debye length scale in the source ohmic region, and somewhat greater in the injection region. This implies that $\gamma_1<1$, while $\gamma_2>1$. Figure 2.1.10a shows the potential calculated from equation (2.1.16), and that calculated from *PC-1D*, a numerically-based Poisson solver [2.1.26]. The x-axis is scaled by λ_{Dch}. It is seen that for $\gamma_1=0.05$ and $\gamma_2=8$, good agreement is obtained over the injection region. The larger discrepancies around $x=-L_{GS}$ result from degenerate effects not included in equation (2.1.16). The concentration in the injection region is obtained from substituting equation (2.1.16) in the Boltzmann distribution.

$$n(x) = N_c \exp\left[-\frac{q}{kT}\psi_{0ch}\left(1 - \alpha_2 e^{-\gamma_2 \frac{x-L_{GS}}{\lambda_{Dch}}}\right)\right] \tag{2.1.19}$$

Noting that $\alpha_2 \approx 1-\psi_{01}/\psi_{02}$ when $\lambda_{DS} << \lambda_{Dch}$, a simplified expression can be obtained when $x<<\lambda_{Dch}$:

$$n(x) \approx \frac{N_{DS}N_{Dch}}{N_c} \exp\left[-\frac{q}{kT}(\psi_{0ch} - \psi_{0S})\cdot\gamma_2 \frac{x - L_{GS}}{\lambda_{Dch}}\right] \tag{2.1.20}$$

Figure 2.1.10b shows the concentration as calculated from equation (2.1.20) compared to that calculated using *PC-1D*. The agreement is good over most of the injection region.

The next region to discuss is that beneath the gate ($0<x<L_G$). Under normal operation, this area is depleted by the built-in voltage of the Schottky contact. This forms the basis to derive the threshold voltage V_{th} by integrating Poisson's equation. Variation of the potential in

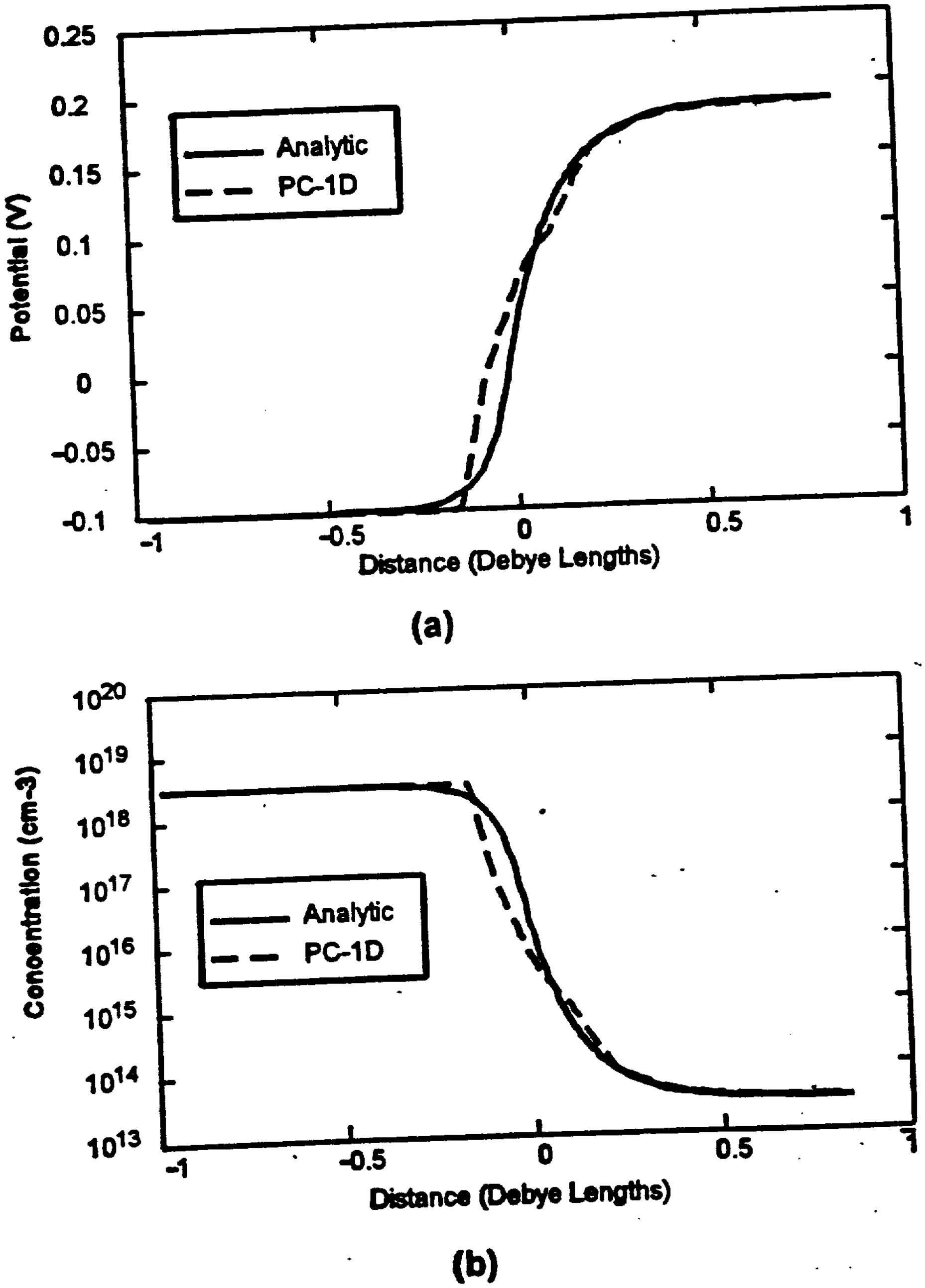

Figure 2.1.10 (a) Injection region potential calculated analytically and from PC-1D. (b) Electron concentration calculated analytically and from PC-1D.

the x-direction can be neglected if the gate is longer than a few Debye lengths, and the integration done in one dimension. In the absence of a gate bias, the total potential in this region is the superposition of the Schottky barrier, depletion, and conduction bandgap potentials, provided they exist. The threshold voltage then is given by the potential at the bottom of the depletion region, which is assumed to extend to the bottom of the channel layer. In the case of the p-HACT structure proposed by Smith, *et al.* [2.1.10], a *pn* junction is used to form the top barrier, and the gate is recessed to the charge control layer. The threshold voltage for this is given by

$$V_{th} = q\phi_b + \frac{q}{\epsilon_s}\left(d_{cc}^2 N_{Dcc} + d_{ch}^2 N_{Dch}\right) \tag{2.1.21}$$

where $q\phi_b$ is the Schottky barrier height, d_{cc} is the thickness of the charge control layer, and N_{Dcc} its donor concentration, and d_{ch} is the thickness of the transport channel layer with doping $N_{Dch.}$ In the case where an AlGaAs layer is used to form the barrier at the top of the channel, the conduction band discontinuity ΔE_c must be included in the calculation of threshold voltage. The gate is not usually recessed in this material structure. Thus V_{th} is given by

$$V_{th} = q(\phi_b - \Delta E_c) + \frac{q}{\epsilon_s}\left(d_{cp}^2 N_{Dcp} + d_{cc}^2 N_{Dcc} + d_{ch}^2 N_{Dch}\right) \tag{2.1.22}$$

where d_{cp} is the cap layer thickness, and N_{Dcp} is its doping.

The last region to consider is the channel region, $x>L_G$. The equilibrium concentration is obviously N_{Dch} at a location several Debye lengths from the end of the gate. Being many orders of magnitude smaller than N_{DS}, this creates a barrier in equilibrium ϕ_w, which must be overcome by the SAW potential. The height of this *injection barrier* is given by

$$\phi_i = \frac{kT}{q}\left[\ln\left(\frac{N_c}{N_{Dch}}\right) + F_{1/2}^{-1}\left(\frac{N_{DS}}{N_c}\right)\right] \tag{2.1.23}$$

where $F_{1/2}^{-1}$ is the inverse of the Fermi integral function

$$F_{1/2}(\xi_f) = \int_0^{\infty} \frac{\xi^{1/2}}{1+e^{\xi-\xi_f}}\, d\xi \quad . \tag{2.1.24}$$

This is not an easily invertible function; however the tabulated values indicate that the conduction energy lies about $6kT$ below the degenerate level for doping on the order of 10^{19} cm^{-3} [2.1.19]. As a minimum, the SAW potential amplitude ϕ_{s0} must exceed the injection barrier (~0.35 V for N_{Dch}=10^{14} cm^{-3}), or no net current will flow into the channel. The injection barrier is plotted as an operational boundary in Figure 2.1.7 in Section 2.1.2.

2.1.3.2 Nonequilibrium Carrier Concentration.

When a SAW potential $\phi_s(x, t)$ is propagating through the HACT mesa, the concentration in the source and injection regions are only perturbed a small amount since the polarization charge concentration is many orders of magnitude smaller. Hence, the effects of the SAW potential are neglected here. As seen from the quasi-equilibrium ACT model developed in section 2.1.1, the concentration in the channel region is greatly affected. A similar situation occurs in the gate region. When a gate bias voltage V_{GS} is connected between the source and gate contacts, a depletion depth changes beneath the gate. The total potential beneath the gate is now given by V_{GS} - V_{th} + $\phi_s(x, t)$, assuming negligible current flow into the gate and negligible screening from any mobile carriers that might be present. The free electron concentration at the input of the gate region, $n(x=0,t)$ is then given by

$$n(0,t) = \frac{N_{DS}N_{Dch}}{N_c}\exp\left[-\frac{q}{kT}(\psi_{0ch}-\psi_{0S})\gamma_2\frac{-L_{GS}}{\lambda_{Dch}}\right] \cdot\exp\left[-\frac{q}{kT}\left[V_{GS}-V_{th}+\phi_s(L_G,t)\right]\right] \tag{2.1.25}$$

assuming that L_{GS} is small enough that the concentration is much greater than the channel doping N_{Dch} and the acoustic polarization charge N_{s0}. It is seen that the gate voltage controls the amount of charge available for transport by the SAW potential, if the conditions specified in section 2.1.1 are met.

2.1.3.3 Drift-Diffusion Injection Theory.

Having established the concentration in the injection region, the next step toward formulating the injection model is to analyze the motion of the electrons in the presence of the SAW potential. The drift-diffusion model under steady state assumptions is adequate to describe the motion of carriers in the injection region because of the slow variation of the potential in the transport direction [2.1.5]. When a SAW potential with an amplitude in excess of the injection barrier is propagating, channel current is expected to flow from source to drain due to the carrier bunching effects discussed in the last section. The application of a negative gate bias will limit this current by deepening the gate depletion region, and thus heightening the barrier over which the carriers must flow. A common technique used to evaluate carrier transport over a barrier is to integrate the current over the barrier region. Since the current is constant, it can then be evaluated from the boundary conditions at the depletion region edges [2.1.14]. Assuming that the carrier concentration in the GaAs transport channel is uniform in the y-direction, the instantaneous channel current I_c through the gate region is

$$I_c(x,t) = -Wd_{ch}q\left(n(x,t)\,\mu_n\frac{d\phi_s}{dx} + D_n\frac{dn}{dx}\right) \tag{2.1.26}$$

where W is the channel width, μ_n is the electron mobility, and D_n is the diffusion constant for electrons. In keeping with the development of the MESFET subthreshold model of Liang, *et al.*, the drift current term in equation (2.1.26) is neglected because of the low concentration of carriers in the depletion region, and because the potential beneath the gate is relatively flat [2.1.22]. Assuming that recombination in the depletion region can be neglected, the current continuity equation forces the current to be constant throughout this region. Hence, integration of equation (2.1.26) results in a simple expression for the injected channel current:

$$I_c(t) = -\frac{qWd_{ch}D_n}{L_G}\left[n(L_G,t) - n(0,t)\right] \ . \tag{2.1.27}$$

An expression for the concentration at x=0 is given in equation (2.1.25). Initially, the well into which charge is injected into is empty, and $n(L_G,0) << n(0,0)$. Hence, the channel current can be approximated by

$$\begin{aligned} I_c \approx -\frac{Wd_{ch}q\ D_n}{L_G}\cdot\frac{N_{DS}N_{Dch}}{N_c}\cdot\exp\left[-\frac{q}{kT}(\psi_{0ch}-\psi_{0S})\cdot\gamma_2\frac{x-L_{GS}}{\lambda_{Dch}}\right] \\ \cdot\exp\left[\frac{q}{kT}\left(V_{GS}-V_{th}+\phi_s(x,t)\right)\right]. \end{aligned} \tag{2.1.28}$$

In reality, the well is filling up as charge is injected, hence decreasing the rate of charge diffusion. I_c will decrease substantially from this value as the charge packet concentration increases to a significant fraction of $n(0,t)$. Hence, equation (2.1.28) is valid in the low injection limit. The rate equation can be solved to obtain a more exact form of the injected charge. However, this rate equation is nonlinear because concentration in the well changes as a function of the charge load. In order to keep a closed-form expression, a simple empirical modification to equation (2.1.28) can be made by introducing an ideality factor η:

$$\begin{aligned} I_c(V_{GS},t) = -\frac{W d_{ch} q D_n}{L_G}\frac{N_{DS}N_{Dch}}{N_c}\cdot\exp\left[-\frac{q}{kT}(\psi_{0ch}-\psi_{0S})\cdot\gamma_2\frac{x-L_{GS}}{\lambda_{Dch}}\right] \\ \cdot\exp\left[\frac{q}{\eta kT}\left(V_{GS}-V_{th}+\phi_s(0,t)\right)\right]. \end{aligned} \tag{2.1.29}$$

Strictly speaking, the ideality factor η is a function of gate voltage V_{GS} through its relation to the well charge. However, proper selection of this parameter results in agreement over wide ranges of operation. It should be pointed out that some ideality factor would have to be introduced into equation (2.1.28) regardless of the well filling effect, to account for recombination in the depletion region. It has been found for depletion regions of lower doped *pn* junctions that $\eta=2$ results in a good fit to the measured *I-V* characteristics because of the bulk recombination effect [44]. So it is expected that $\eta=2$ would be the asymptotic limit for low injection, and that η would increase as the average injection current increased.

2.1.3.4 Gaussian Approximation for the Instantaneous Channel Current.

The expression given in equation (2.1.29) for the instantaneous channel current is closed-form, but is not easily integrable because of the cosine term in the exponential. For the purposes of calculating the steady-state channel current, the cosine is expanded out to two terms in a Taylor series. A sampling aperture function *S(t)* can then be defined which is essentially Gaussian in form,

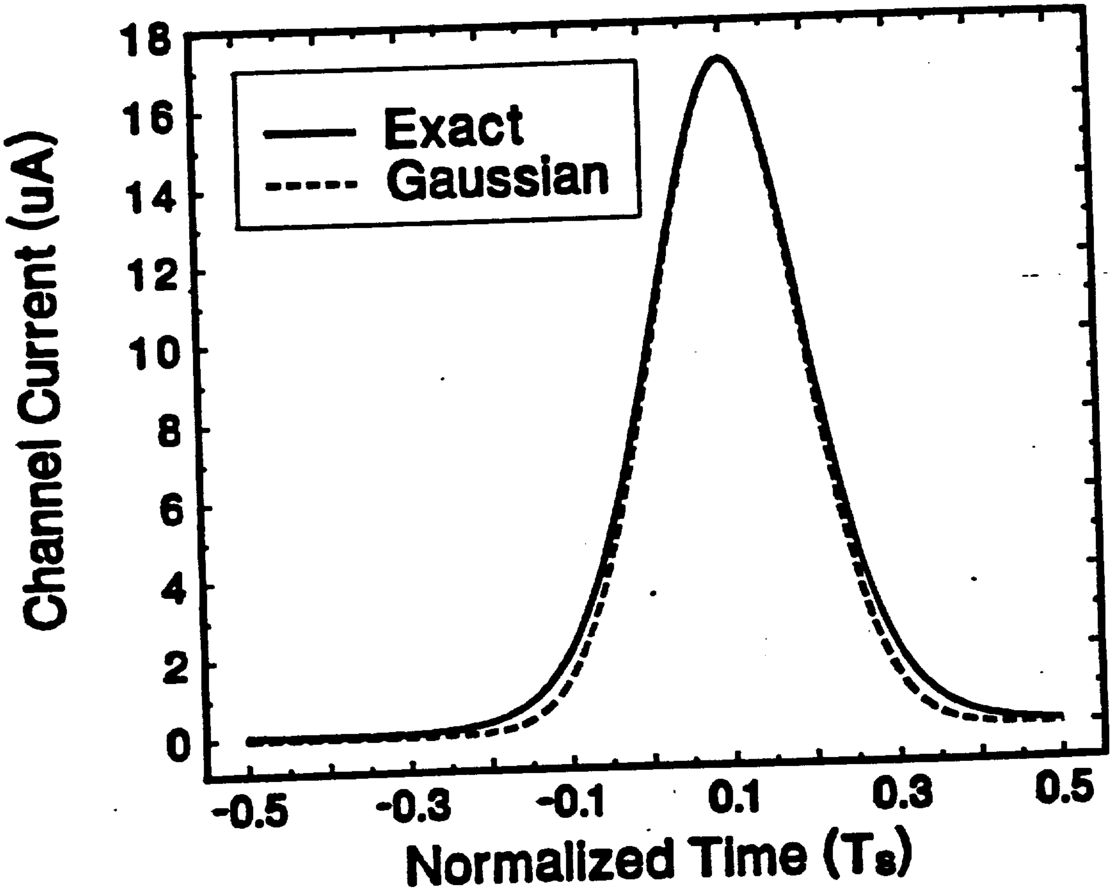

Figure 2.1.11 Calculated sampling aperture function using exact formulation and Gaussian approximation for a 1 mm wide channel, $\eta = 5.5$ and $\phi_s = 0.26$ V.

$$
\begin{aligned}
I_c(V_{GS},\phi_{s0},t) \approx & -\frac{W d_{ch} q D_n N_{DS}}{L_G}\cdot\exp\left(-\frac{L_{GS}}{\lambda_D}\right) \\
& \cdot\exp\left(\frac{q}{\eta k T}(V_{GS}-V_{th}-\phi_{s0})\right)\cdot S(t) * \sum_{i=-\infty}^{\infty}\delta\left(t-\frac{i}{2\pi f_s}\right)
\end{aligned}
\qquad (2.1.30)
$$

where

$$
S(t) = e^{-\frac{t^2}{2\sigma^2}}, \qquad \sigma = \frac{1}{\pi f_s}\sqrt{\frac{\eta k T}{q\phi_{s0}}}\,. \qquad (2.1.31)
$$

Plots of $I_c(t)$ calculated from Eqs. (2.24) and (2.25) are shown in Figure 2.1.11. It is seen that there is little error introduced by the Gaussian approximation. Defining the sampling width to be σ as given above would imply that 67% of the charge injected into the well occurs during this time. The approximate width of the sampling aperture has been the subject of much debate. It has been claimed that the aperture is a very small fraction of a SAW period [2.1.6], [2.1.7], [2.1.27]. From equation (2.1.31), typical values for HACT device designs reported in [2.1.7] result in apertures approximately 30% of the SAW period. This is experimentally verified in the next section.

2.1.3.5 Average Channel Current.

The average channel current $\langle I_c(V_{GS})\rangle$ can be obtained by integrating over one SAW period T_s. This can be done by numerically integrating equation (2.1.29), or analytically by expanding the limits of integration to infinity on the Gaussian definite integral formed from equation (2.1.30). In doing the later, an approximate analytical expression for the average channel current is developed:

$$
\begin{aligned}
\langle I_c(V_{GS},\phi_{s0})\rangle &= \int_{-T_s/2}^{T_s/2} I_c\left(V_{GS},\phi_{s0}\cos(2\pi f_s t)\right) dt \\
&\approx -\frac{W d_{ch} q\, D_n N_{DS}}{L_G}\cdot\exp\left(-\frac{L_{GS}}{\lambda_D}\right) \qquad (2.1.32) \\
&\cdot\frac{1}{\pi f_s}\cdot\sqrt{\frac{\eta k T}{q\phi_{s0}}}\cdot\exp\left[\frac{q}{\eta k T}(V_{GS} - V_{th} + \phi_{s0})\right].
\end{aligned}
$$

equation (2.1.32) predicts an exponential dependence of the average channel current on the gate voltage, which is the same as observed in MESFET's and HEMT's in the subthreshold region. This is verified experimentally in the next section.

2.1.4 Experimental Technique for Injection Model Verification.

Verification of the analytical injection model presented in this chapter was made by measurement of HACT devices fabricated at Georgia Tech and United Technologies, Inc. The two salient features predicted by the injection model are the Gaussian shape of the sampling aperture and the exponential dependence of the average channel current on the gate voltage.

2.1.4.1 Injection Aperture Test Setups and Procedures.

The Gaussian shape of the aperture function has been verified by measuring the conversion loss of the input sampler with input signal frequencies at harmonics of the Nyquist frequency. The mathematical basis of this measurement, and how measurement of the conversion loss relates to the aperture shape, can be understood by referring to Figure 2.1.12. This shows the functional block diagram of the HACT devices on which these measurements were performed. The output current is the product of the sampling function and the input voltage signal. Since

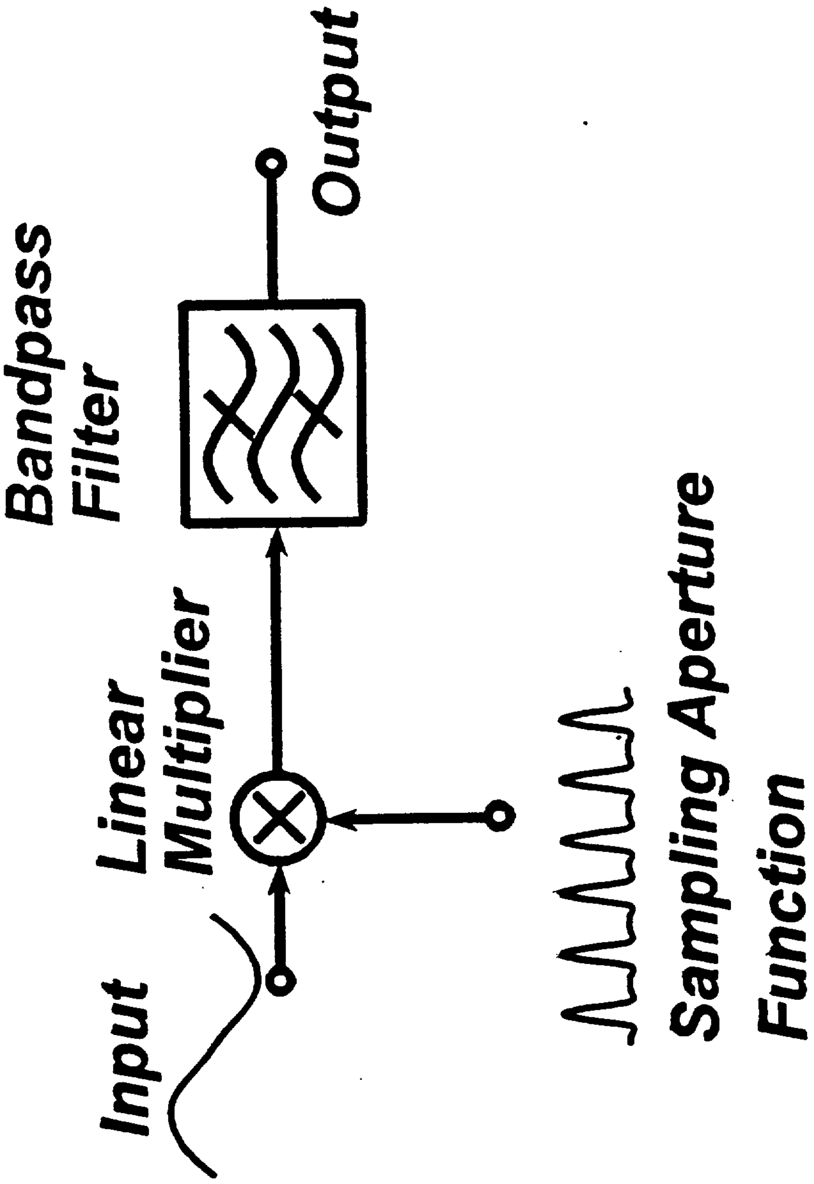
Input
Linear
Multiplier
Bandpass
Filter
Output
Sampling Aperture
Function

Figure 2.1.12 System level representation of the HACT device for aperture function measurement.

the sampling function is periodic, it can be expressed as a Fourier series

$$S(t) = \sum_{m=-\infty}^{\infty} a_m e^{jm\omega_0 t} \tag{2.1.33}$$

where $j=(-1)^{1/2}$, $\omega_0 = 2\pi f_s$, and a_m is the *mth* Fourier coefficient. Thus, the determination of these coefficients leads to the shape of the sampling aperture. Assuming that $S(t)$ is real and even, then $a_m \geq 0$ for all m. If a sinusoidal signal, $v_k(t)$, of a frequency some integer multiple of the Nyquist frequency

$$v_k(t) = \cos\left(\frac{k\omega_0}{2}\right) \tag{2.1.34}$$

is applied to the gate, then the output channel current, bandlimitted by the bandpass filter function of the NDS array, i_k can be expressed as

$$\tilde{i}_k(t) = \frac{1}{2} a_{m-k} \cos\left(\frac{2m-k}{2}\omega_0 t\right) \delta_{2m,k-1} \tag{2.1.35}$$

where δ is the Kronecker delta function. It is seen that the amplitude of i_k is twice the value of the Fourier coefficient a_{m-k}. Thus, measurement of the power at the *kth* multiple of the Nyquist frequency will give the Fourier coefficient, and hence the shape of the injection aperture.

2.1.4.2 Injection Aperture Measurements and Results.

The injection aperture was measured on two HACT devices with identical input structures using the setup shown in Figure 2.1.13. The time-domain waveform was reconstructed from measurement of the first three Fourier coefficients. The resulting aperture function as calculated using equation (2.1.33) is shown in Figure 2.1.14, along with the Fourier reconstructed sampling function.

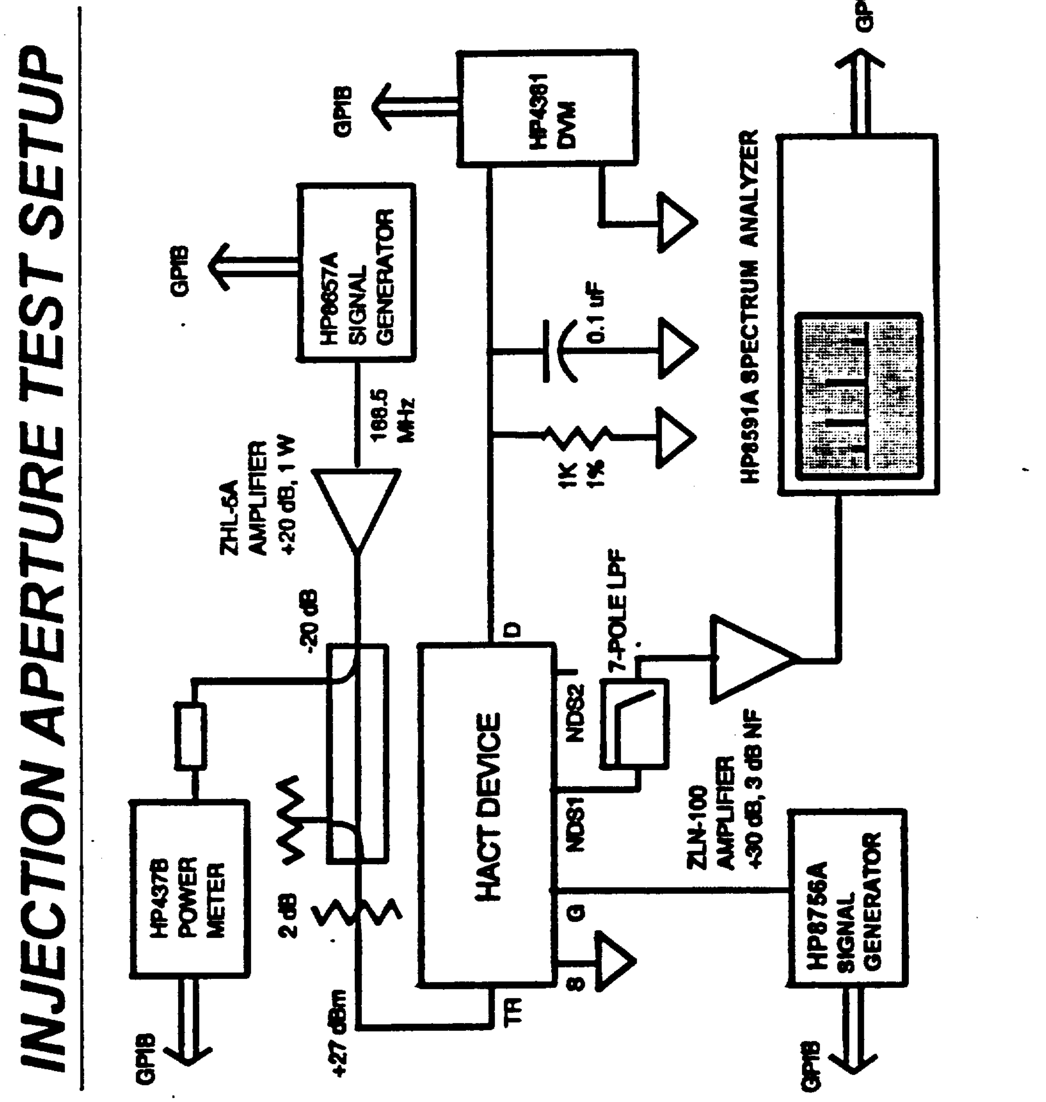
INJECTION APERTURE TEST SETUP
GPIB
HP437B POWER METER
2 dB
-20 dB
+27 dBm
TR
HACT DEVICE
S
G
NDS1
NDS2
D
ZHL-6A AMPLIFIER +20 dB, 1 W
166.5 MHz
HP8657A SIGNAL GENERATOR
GPIB
HP4361 DVM
GPIB
0.1 uF
1K 1%
7-POLE LPF
ZLN-100 AMPLIFIER +30 dB, 3 dB NF
HP8591A SPECTRUM ANALYZER
GPIB
HP8756A SIGNAL GENERATOR
GPIB

Figure 2.1.13 Test setup used to measure injection aperture shape. Input signals are multiples of the Nyquist frequency. Output is input frequency translated by a harmonic of the aperture function.

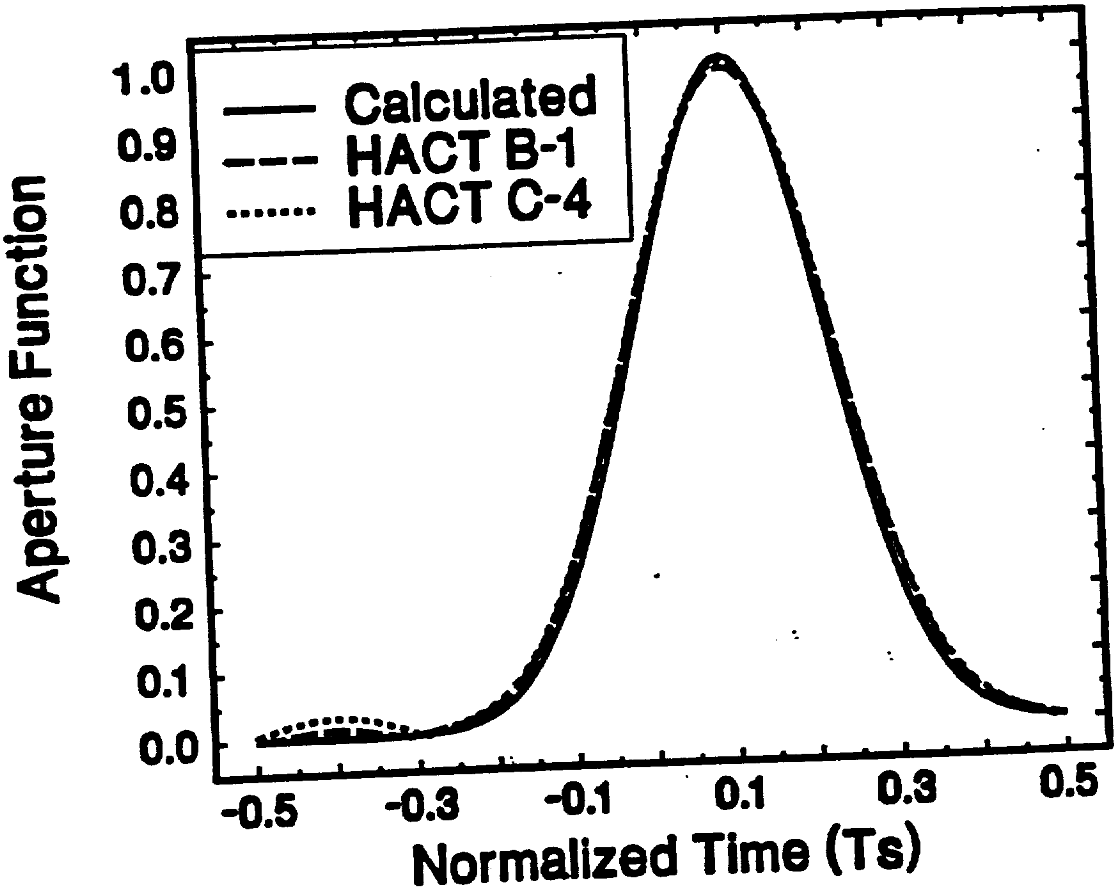

Figure 2.1.14 Measured versus predicted aperture function for two 1 mm wide HACT devices.

2.1.4.3 Average Channel Current Test Setups and Procedures.

The relationship between steady-state channel current and gate voltage was also verified. This was done using the test setup discussed in section 2.1.2. The average channel current is calculated from the approximate Gaussian integral given in equation (2.1.32), and by numerically integrating equation (2.1.30). The difference between these is too small to discern, hence both results appear as the "predicted" values in Figure 2.1.15. The measured $\langle I_c \rangle$-V_{GS} curve of a 1 mm wide HACT device designed and fabricated by United Technologies is also shown in Figure 2.1.15. The ideality factor η and was adjusted so that the calculated $\langle I_c \rangle$ was in close agreement with the measured average channel current. It was found that η=3.3 gave good results over a wide range of channel current values.

2.1.4.4 Average Channel Current Measurements and Results.

The nature of the average channel current as a function of gate-source voltage has been the subject of some debate. An exponential dependence was shown by the numerical model developed by Bogus for buried channel ACT devices [2.1.1]. However, Merrit observed a square-law dependence in HACT devices, similar to what is seen in a MESFET above threshold [2.1.8]. It was observed in this research that the measured channel current does follow an exponential dependence as predicted by the injection model for low injection levels. However, the variation of the ideality factor with gate voltage causes a less than exponential dependence in the $\langle I_c \rangle$-V_{GS} characteristic. It is possible that this variation would result in an empirical fit to a square-law characteristic. However, the physical basis for this is due to completely different mechanisms than are present in a MESFET operating in the saturated region.

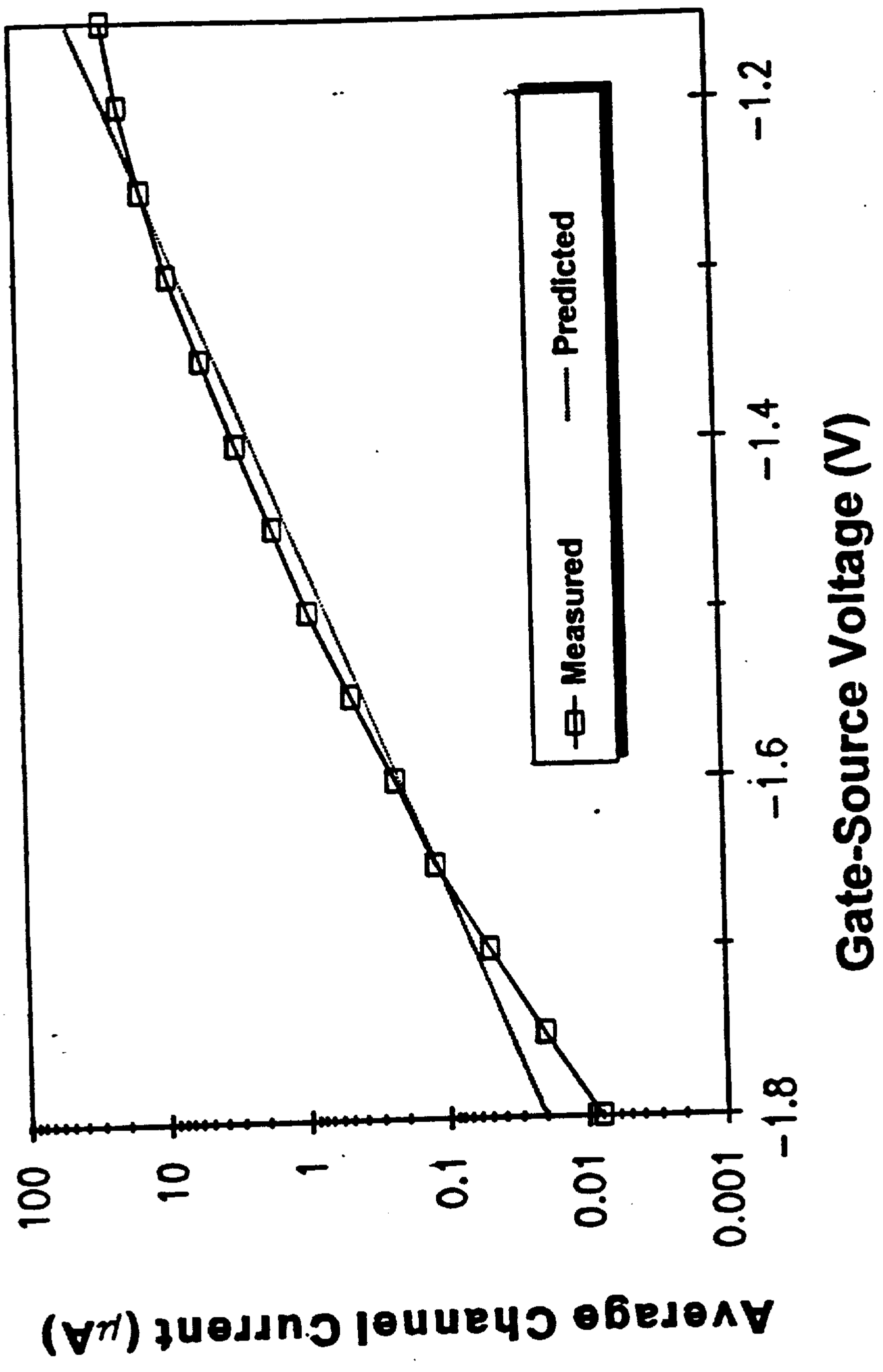

Figure 2.1.15 Measured versus predicted $<I_c>$ plotted against gate bias voltage for a 1 mm wide HACT device, $\eta = 3.3$ and $\phi_s = 0.26$ V.

2.1.5 Summary of Results and Conclusions

This chapter has presented the first analytic theory of charge transport and injection in HACT devices developed from the physics of the device operation. The transport model includes expressions for charge packet concentration, charge capacity, and barrier height between adjacent packets, in terms of the physical device parameters and operating conditions. It was seen that charge transport is greatly affected by channel doping, and there exist maximum and minimum conditions for both the doping and SAW potential in order for charge transport to occur. The theory was verified against measured data for a number of HACT devices fabricated at Georgia Tech and elsewhere. The injection theory predicts the instantaneous and average channel current as a function of device geometry and operating conditions. This was the first analytic description of the injection aperture shape, which was found to be approximately Gaussian. The physical mechanism of subthreshold current injection was also used to explain the average channel current dependance on gate voltage being close to exponential. Both the instantaneous and time-averaged channel current functions were verified experimentally.

The work presented in this chapter provides a framework within which one can quantify tradeoffs in the design of HACT material architecture and device geometry. The model can also be used to predict the limits of the technology for specific applications.

2.2 Piezoelectric Thin Films on GaAs-Based materials

A potential application for piezoelectric films on GaAs substrates is the monolithic integration of surface acoustic wave (SAW) devices with GaAs electronics. Knowledge of the SAW properties of the layered structure is critical for the optimum and accurate design of such devices. The acoustic properties of ZnO films sputtered on {001}-cut <110>-propagating GaAs substrates are investigated in this article, including SAW velocity, effective piezoelectric coupling constant, propagation loss, diffraction, velocity surface, and reflectivity of shorted and open metallic gratings. The measurements of these essential SAW properties for the frequency range between 180 and 360 MHz have been performed using a knife-edge laser probe for film thicknesses over the range of 1.6 to 4 μm and with films of different grain sizes. The high quality of dc triode sputtered films was observed as evidenced by high K^2 and low attenuation. The measurements of the velocity surface, which directly affects the SAW diffraction, on the bare and metalized ZnO on SiO_2 or Si_3N_4 on {001}-cut GaAs samples are reported using two different techniques: 1) knife-edge laser probe, 2) line-focus-beam scanning acoustic microscope. It was found that near the <110> propagation direction, the focussing SAW property of the bare GaAs changes into a non-focussing one for the layered structure, but a reversed phenomenon exists near the <100> direction. Furthermore, to some extent the diffraction of the substrate can be controlled with the film thickness. The reflectivity of shorted and open gratings are also analyzed and measured. Zero reflectivity is observed for a shorted grating. There is good agreement between the measured data and theoretical values.

2.2.1 Background:

Over the past two decades, ZnO has been the most frequently considered piezoelectric film for use for increasing the piezoelectric coupling in cases of a weakly- (*e.g.* GaAs) or non-piezoelectric substrate [2.2.1]. The use of a ZnO layer with a GaAs substrate will enable one to monolithically integrate surface acoustic wave (SAW) devices with GaAs electronics. Furthermore, such a structure will have low dispersion due to the similarity of the SAW velocities of both materials [2.2.2]. In this paper, the feasibility of ZnO films deposited on GaAs substrates for SAW device applications is investigated along with a detailed discussion about the problems which occurred during the characterization of the ZnO/GaAs structures.

Knowledge of SAW properties of the filmed structure is crucial for the accurate design of SAW devices. Herein, the experimentally determined fundamental SAW properties are reported including velocity, effective piezoelectric coupling constant K^2, attenuation for the 1.6 to 4 μm thicknesses of c-oriented ZnO film over semi-insulating {001}-cut <110>-propagating GaAs substrates in the frequency range of 180-360 MHz.

Another critical acoustic property for the design of SAW devices is diffraction of the surface acoustic wave. Diffraction in anisotropic media can generally be described by the slowness surface or the velocity surface. The velocity surfaces on free and metalized surfaces were measured using two independent techniques: 1) knife-edge laser probe, 2) line-focus-beam scanning acoustic microscope (LFBSAM). The results from both methods are compared with the theoretical results.

Metallic gratings are basic elements required for the construction of SAW devices. Analyzing the reflectivity and the velocity change due to metallic gratings has been studied since the invention of SAW devices; however, no research on gratings on multilayered structures has been reported. Datta and Hunsinger have analyzed the grating properties on a single crystal by separating into the piezoelectric shorting and the mechanical scattering due to the mass loading and the stress [2.2.3]. In this paper, an extension of their technique to a multi-layered structure is presented. Experimental results for a shorted and an open grating are compared with theoretical calculations.

A passivation layer such as SiO_2 [2.2.4], Si_3N_4 [2.2.5], or SiON [2.2.6] between ZnO and GaAs might be required in order to: 1) passivate the structure and enhance the yield, and 2) prevent unwanted doping of GaAs by Zn during the sputtering processes. The application of such a passivation layer was obtained with a plasma-enhanced chemical vapor deposition (PECVD) for SiO_2 and CVD for Si_3N_4. The acoustic parameter differences with and without the passivation layer are discussed. Theoretical calculations of SAW velocity and K^2 as a function of film thickness were performed using the Laguerre polynomial technique [2.2.7].

2.2.2 ZnO film growth

C-axis oriented polycrystalline ZnO films were grown at Motorola using both rf magnetron and dc triode sputtering method on {001}-cut GaAs wafers which were grown by the liquid-encapsulated Czochralski (LEC) technique. The GaAs wafers, which were chemically and mechanically polished on both sides, were semi-insulating with resistivity $\sigma = 10^7 \Omega$-cm, and the surface normal direction was [100] ± .1°. The quality of the film strongly depends on the fabrication conditions. The typical deposition parameters used for the ZnO film are listed in Table 2.2.1. The film thicknesses of 1.6, 2.8, and 4.0 μm have been chosen to cover the range of 0.1-0.5 of the acoustic wavelength, λ. As a passivation layer, a 1000Å PECVD SiO_2 layer was deposited for the rf magnetron sputtered ZnO films and a 2000Å CVD Si_3N_4 layer was deposited for the dc triode sputtered films. The choice of a passivation layer was made solely on the equipment availability.

The as-grown ZnO films were transparent with a very smooth surface finish. The grain size of the ZnO films was 0.2 to 0.5 μm for the rf magnetron sputtered films, and no grain boundaries are visible under scanning electron microscope (SEM) examination for the dc triode sputtered films indicating the grain size was beyond the resolving power of the SEM. From X-ray diffraction and atomic force microscope (AFM) measurements, the grain size of the dc triode sputtered films is 30 to 50 nm. The problem with these particular dc triode sputtered films, however, is that a compressive stress is induced in the substrate at the time the film is grown. The induced stress is greater with the thicker films such that 2" wafers are warped, making it difficult to process IDTs. Thus, only the ZnO film thickness of 1.6 μm was investigated for the dc triode sputtered films. It has been previously reported that a decrease in the internal compressive stress results in a decrease in K^2 [2.2.8].

The orientation of the ZnO film was characterized using the X-ray diffraction technique. The diffraction patterns showed that all samples had strong c-axis texture; *i.e.*, there were no ZnO peaks visible except those from the basal planes.

2.2.3 Characterization of acoustic properties

A number of devices were fabricated using a standard lift-off process on 1.6, 2.8, and 4 μm thicknesses of the ZnO films. Employing one set of four IDTs with different center frequencies enables the investigation of frequency characteristics over the range of 180-360 MHz. The configuration of the IDTs is listed in Table 2.2.2. The values of the SAW wavelength, λ, in the table denote nominal values. Each IDT is unanodized with metallization ratio (finger width to grating period) of 0.6. The apertures of the IDTs were chosen to facilitate the measurement of the slowness surface on bare GaAs substrates [2.2.9]. Considering the non-acoustic resistivity of the IDTs [2.2.10], the metal thickness was chosen to be 1000 Å of Al/4% Cu alloy. For ease of fabrication, the smallest dimension for the devices was no less than 2 μm; therefore, IDT No.1 utilized split finger electrodes while the others utilized the single finger configuration. Acoustic absorber was applied at both ends of the device in order to prevent SAW reflections from the edges.

2.2.3.1 Vsaw

The value of λ is inferred from the periodicity of the IDT electrodes, and the SAW velocity, V_{saw}, in the IDT region may be obtained to first order by the relation, $V_{saw} = f_c/\lambda$, where f_c represents the center frequency. The value of the nominal λ within the free ZnO/GaAs surface region is different and is measured directly by using a technique such as the laser probe system. The measured values of f_c are shown in the Table 2.2.2, and the variation of f_c is due to the fact that V_{saw} is dispersive depending upon the ZnO film thickness as well as the passivation layer.

A knife-edge laser probe system was used to measure SAW properties. The use of the laser probe system and the scheme for the longitudinal and transverse scans has been described in [2.2.11] and [2.2.12] and will not be discussed here. In this study, the signal to noise ratio of the laser probe was typically 40-50 dB, and a 3 μm diameter laser spot was used. From several devices on each substrate, the best device was chosen for the experiments. The SAW velocity measured with the knife-edge laser probe is shown in Fig. 2.2.1 along with the theoretical values. The velocity dispersion is plotted as a function of wavelength-normalized film thickness. The theoretical values were calculated with the value of λ set to 10 μm. The material constants in the calculation were those given by Slobodnik [2.2.13] and Hickernell [2.2.5].

The maximum difference between the SAW velocities of the free surface (measured with the laser probe) and in the IDT area, from the center frequency measurements, was 0.3%. The maximum velocity deviations caused by the passivation layer were 0.6 and 1.6 % for SiO_2 and Si_3N_4, respectively, and these figures agree well with the values from the center frequency measurements. The accuracy of the velocity measurement is approximately 0.3%. The detailed discussion about the accuracy can be found in Bright, Kim, and Hunt [2.2.14] and will not be reproduced here.

In using the laser probe, difficulty in making the free surface measurements was experienced due to the multiple reflections of the laser beam between the film-substrate interface and the air-film. For three film structures (2.8 and 4.0 μm rf magnetron sputtered ZnO and 1.6 μm dc triode sputtered ZnO/0.2μm Si_3N_4), no appreciable energy from the laser beam was reflected from the free surface areas which is similar to what happens with

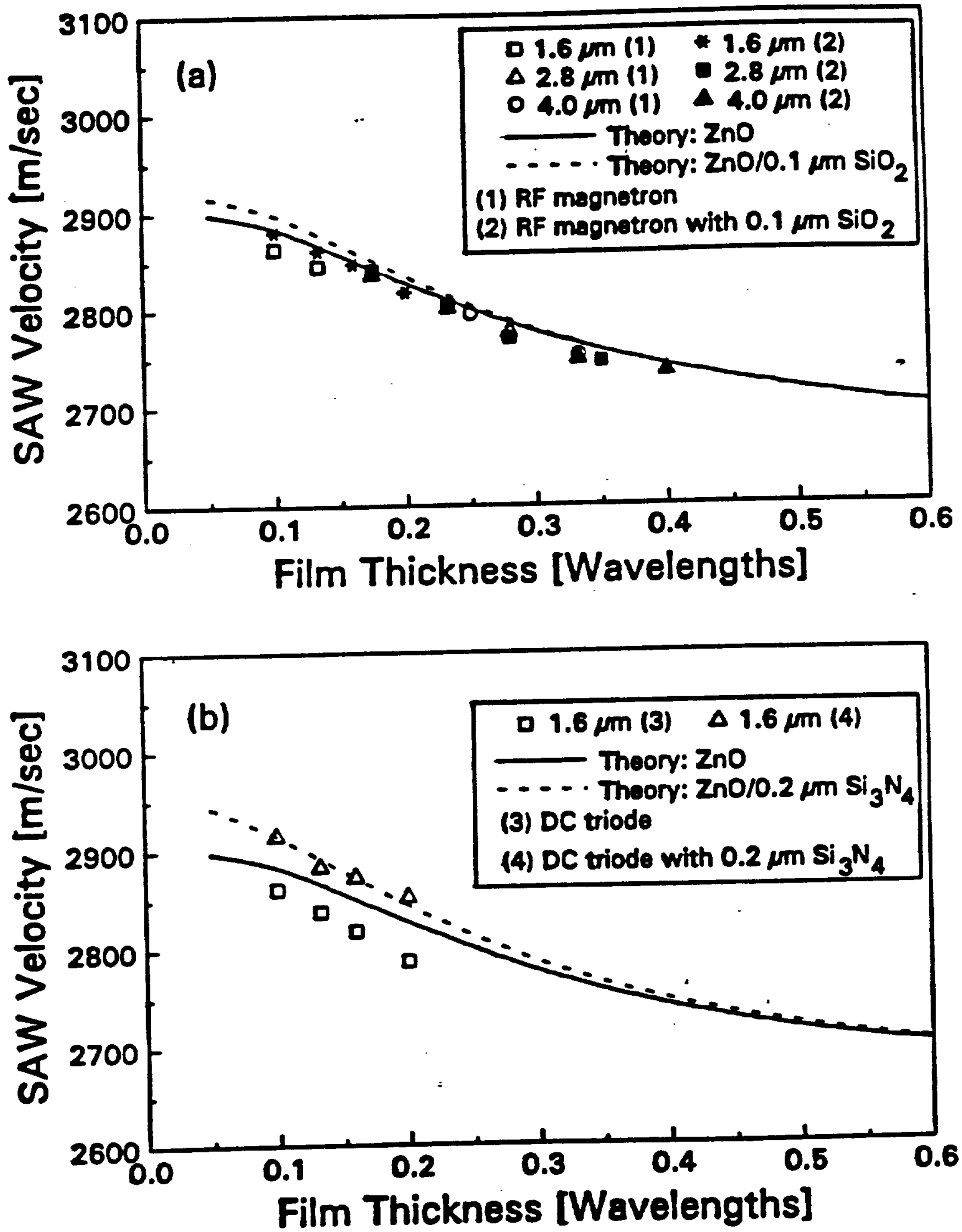

Figure 2.2.1 SAW velocity measured by a knife-edge laser probe on the metalized surface:
(a) RF magnetron sputtered film,
(b) DC triode sputtered film.

anti-reflection coatings in optics. This made probing impossible. For this reason, the velocity shown in Fig. 2.2.1 is the measured data on the metalized surface as well as the theoretical value on short-circuit surface. More detailed analysis about this multiple reflection can be found in [2.2.15].

2.2.3.2 Piezoelectric Coupling (K^2)

The effective piezoelectric coupling constant K^2 is generally obtained by a measurement of the relative shift in velocity between open- and massless short-circuit surface as follows,

$$K^2 = 2\frac{(V_o - V_s)}{V_o} \tag{2.2.1}$$

where V_o and V_s are the open- and short-circuit velocity, respectively. For the measurement of V_s, a metal pad was fabricated on one side of the IDTs to short out the electric field on the surface while the other side of the IDT was left as a free surface. In order to reduce the mass loading of the metal pad, it was fabricated with 250Å thickness of Al/4% Cu alloy. In spite of the finite thickness of the pad, its mass loading would be negligible because its thickness is only 0.0016 to 0.0031 of λ (16 to 8 μm) and it has a very slight effect compared to the piezoelectric effect [2.2.11]. This assumption was confirmed by a theoretical calculation using the Laguerre polynomial technique [2.2.7]. The velocity change due to the mass loading was calculated to be at most 0.015% which is much less than the expected value of K^2 for the ZnO/GaAs structure.

The value of K^2 calculated by using Eq. (2.2.1) is shown in Fig. 2.2.2 as a function of normalized film thickness. The data points represent average values obtained from several

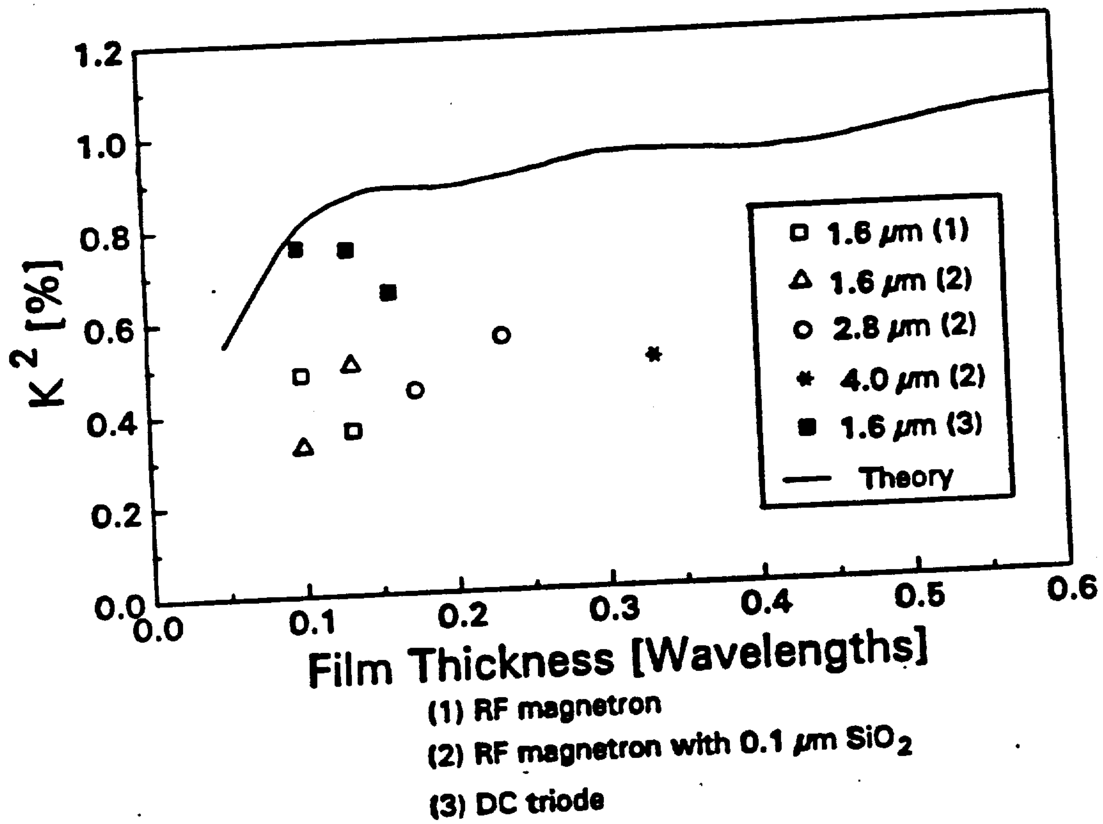

Figure 2.2.2 Piezoelectric coupling constant K^2.

measurements. It is seen that the value of K^2 for the dc triode sputtered film is about 1.7 times larger than that of the rf magnetron sputtered films as expected from the grain size of the film. The measured value is 40-80 % of the theoretical one, which is comparable to Hickernell's observations on the other substrates [2.2.1]. This value is still approximately 5.7-10.8 times larger than the typical value for bare GaAs (0.07 %).

2.2.3.3 Propagation Loss

Acoustic propagation loss (attenuation) is also a crucial parameter in device applications of piezoelectric films. The attenuation of the SAW can be measured by comparing the energy contained in the wave of two different transverse scans, separated by some distance d The energy is proportional to the value obtained by integrating the squared magnitude of the beam profile $|U|^2$ along the transverse direction. The scan length should be made long enough to include the entire energy since the beam spreads out due to diffraction. Thus, the attenuation α is given by

$$\alpha = 10\log\left(\frac{\sum_1 |U|^2\Delta x}{\sum_2 |U|^2\Delta x}\right)\left(\frac{v_o}{d}\right) dB/\mu s \qquad (2.2.2)$$

where Δx is a step size for the scan. It should be noted that this attenuation includes not only 1) so-called viscous loss due to interaction with thermally excited elastic waves, 2) scattering by crystalline defects, impurities, and surface roughness, and 3) air loading but also the losses due to diffraction and the leaky nature of the wave on GaAs [2.2.16]. The effects of loss due to interaction with electrons should be negligible because of the high resistivity of the substrate.

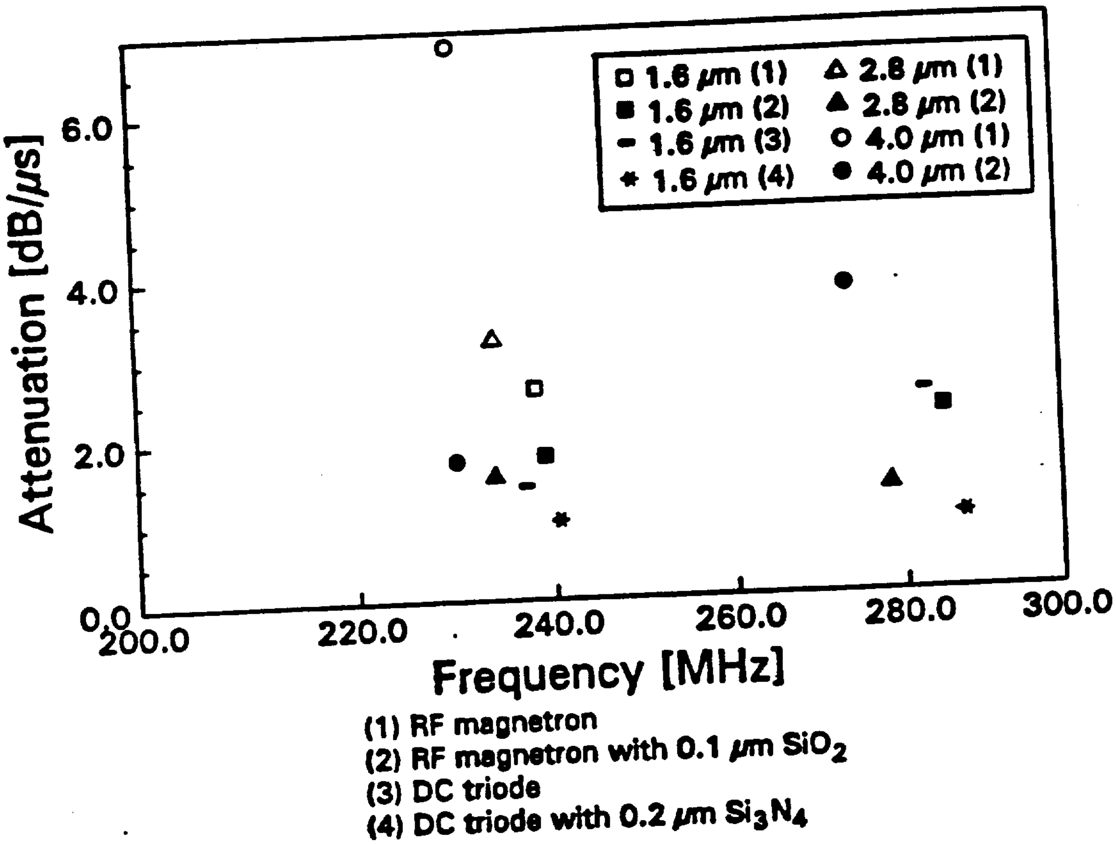

Figure 2.2.3 Propagation loss in the frequency range of 230-290 MHz.

The propagation loss shown in Fig. 2.2.3 was observed in the frequency range of 230-290 MHz, by comparing several transverse scans on the metalized surface with the distance of about 200λ and averaging the measured values. In fact, the diffraction was relatively large compared to that of bare GaAs due to the characteristic of the ZnO film to be discussed in the next section and the narrow aperture of the IDT (W=8.3 or 10λ). With greater diffraction, more SAW energy can be lost to bulk shear since SAW propagation on this cut and propagation direction of GaAs is a leaky SAW mode. Employing an IDT with a wider aperture may reduce the value of the propagation loss in the measurement. Nevertheless, the narrow aperture was used in order to obtain enough diffraction for the velocity surface measurements to be discussed in the next section.

In the laser probe system, the output signal of the laser probe, which is detected with a photo-diode, is filtered into dc and rf components. The dc component signifies the reflectivity of the sample under test, and the rf component is related to the normal component of the SAW particle displacement. While only the phase information of the rf component is required for the velocity measurement, its magnitude is used for the attenuation measurement. Since the magnitude of the rf component is proportional to the dc component, it is normalized by the dc value V_{dc}. The accuracy of the magnitude measurement depends on, among other things, the flatness of the surface. However, the sample is not perfectly flat, resulting in a slight defocus of the laser probe spot and a variation in V_{dc}; thus, precise measurement of attenuation is very difficult. Furthermore, due to the diffraction loss of the IDT, the maximum deviation of the attenuation measurement was up to a few tenths of a dB/µs depending on the scan distance.

The measured data shows that the 1.6 μm dc triode sputtered film with the passivation layer had the lowest loss and the 4 μm rf magnetron sputtered film without the passivation layer the highest attenuation, which is comparable to what has previously been reported [2.2.1],[2.2.17]. The passivation layer had a noticeable effect on both sputtered films-especially for the rf magnetron sputtered ones. The reduction in the attenuation is perhaps due to the fact that the passivation layer provides a better platform for the film growth by compensating for stress differences between the film and the substrate[2.2.2]. In spite of the diffraction loss, it should be noted that the lowest value of attenuation for the films is not severely worse than that of bare GaAs substrate, which has the values of 0.323 and 0.567 dB/μs, for 200 and 300 MHz, respectively [2.2.18].

The ZnO/GaAs structures may not have suitable temperature stability required for some applications due to positive temperature coefficient delays (TCDs) of both bulk ZnO (35 ppm/°C) and GaAs (50 ppm/°C). However, since the passivation layer of SiO_2 (-70ppm/°C) has a strong negative TCD, improvement of the temperature stability may be achieved by adjusting the passivation layer thickness [2.2.2]. An experimental compensation of TCD of GaAs was reported by using a Au/SiO_2 film [2.2.19]. No measurements of the temperature characteristics have been attempted in this research.

2.2.4 Velocity surface

2.2.4.1 Knife-edge laser probe measurements

The substrates investigated for the velocity surface were 1.6, 2.8, and 4.0μm rf magnetron sputtered ZnO films with 0.1μm thick SiO_2 passivation layer and 1.6μm dc triode

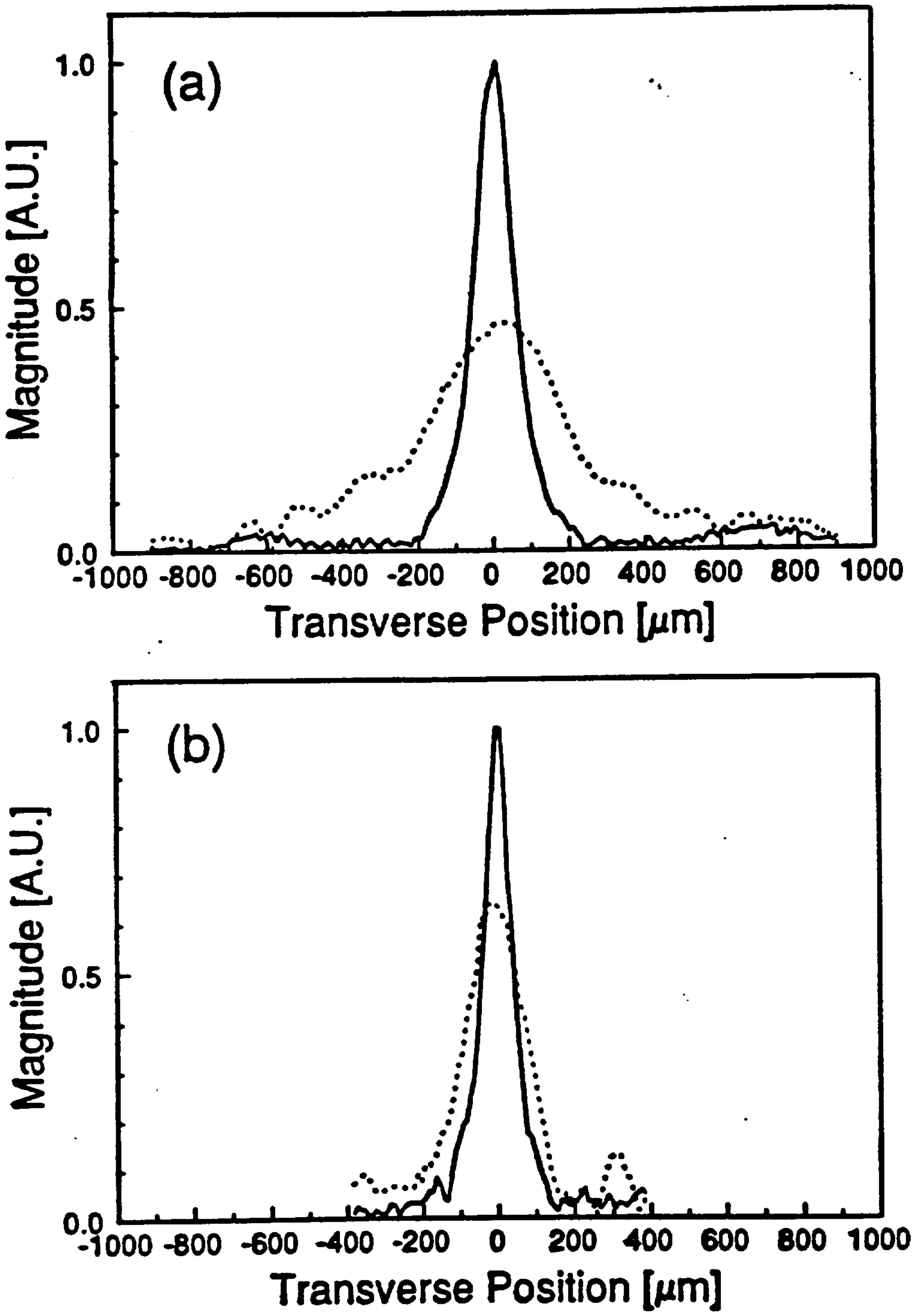

Figure 2.2.4 Beam profiles are separated by 200λ when λ is 12 μm. The solid line is the near the IDT and the dotted line is for 200λ apart. (a) 1.6 μm DC triode sputtered ZnO, (b) bare GaAs.

sputtered ones with and without 0.2μm thick Si_3N_4 passivation when λ was 12 μm. The typical transverse scan data measured on the free surface of 1.6 μm dc triode sputtered ZnO/GaAs and bare GaAs are shown in Figs. 2.2.4(a) and (b), respectively, where the SAW propagation is along the <110> direction of GaAs. The two transverse scans were separated by a distance of 200λ. Note that the acoustic diffraction with the ZnO film shown in Fig. 2.2.4(a) is much larger than that of bare GaAs, which clearly shows an focussing beam profile in Fig. 2.2.4(b). Thus, the velocity surfaces of these structures are investigated theoretically and experimentally to explain the difference in the diffraction.

The theoretical velocity curves of the ZnO film with a 0.1 μm SiO_2 passivation layer are shown in Fig. 2.2.5(a) as a function of the ZnO film thickness normalized to λ. The propagation directions are centered about the <110> axis. Note that the SAW velocity curves for the ZnO film thickness of 0.13, 0.23, and 0.33 λ on GaAs are concave upward, and this is in contrast with that of the concave downward curvature for bare GaAs. The concave downward and upward curvature infer respectively focussing and non-focussing behavior, and support the diffraction shown in Fig. 2.2.4.

The effect of the passivation layer with 0.1μm thick SiO_2 or 0.2μm thick Si_3N_4 is shown along with non-passivated structures in Fig. 2.2.5(b) for the ZnO thickness of 1.6 μm. Since the passivation layer is treated as an isotropic media like the c-oriented ZnO film, its presence is not expected to affect the shape of the curve, but shifts the entire curve slightly toward faster values. Considering the fast SAW velocities of single SiO_2 and Si_3N_4, one can easily expect that the curve would shift upward as the thickness of the passivation layer increases as long as the energy is trapped at the surface.

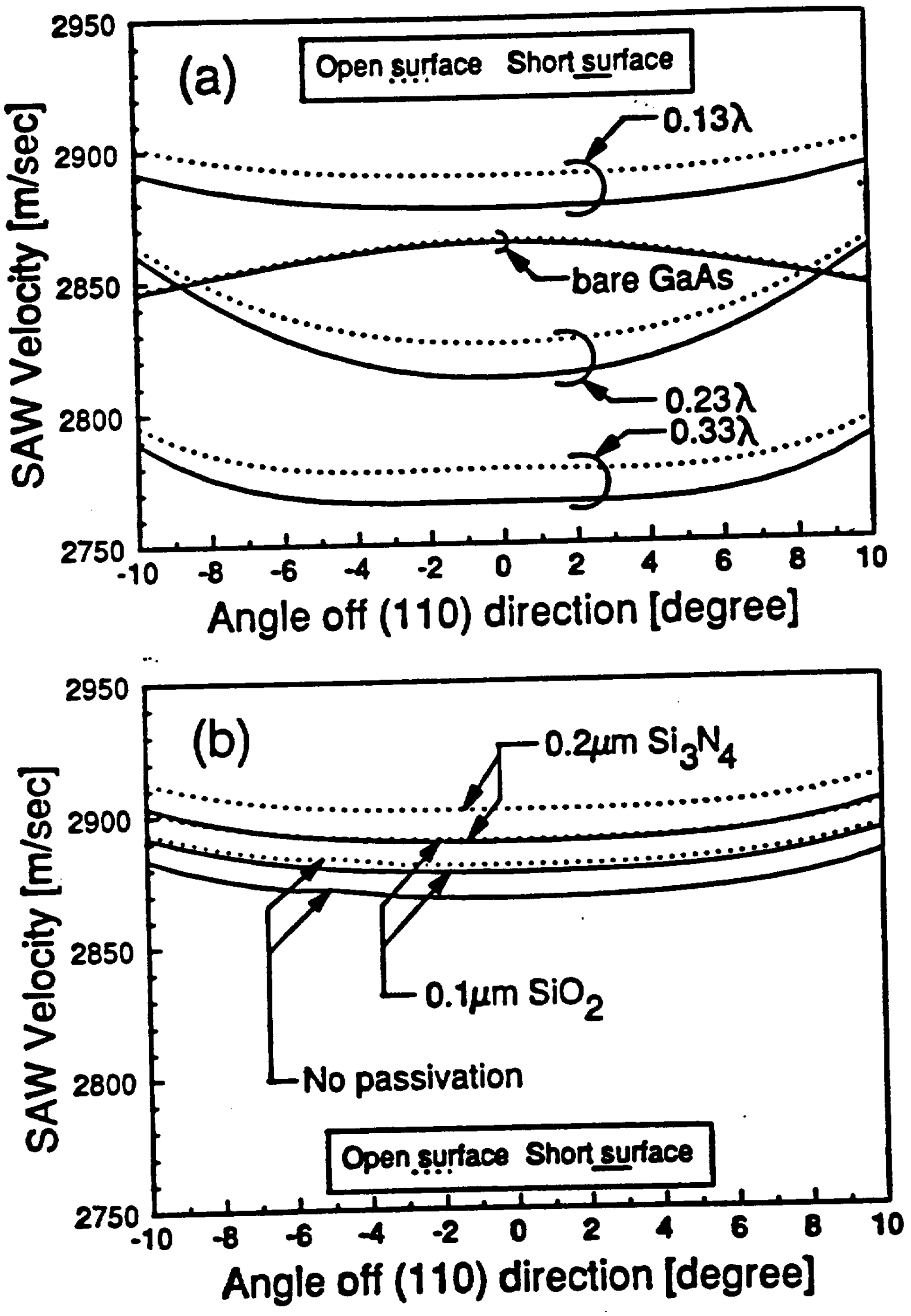

Figure 2.2.5 Velocity surfaces by Laguerre technique. Values indicate thicknesses of ZnO normalized by λ= 12 μm. (a) ZnO film thickness with 0.1 μm SiO_2 layer, (b) 0.13λ thick ZnO.

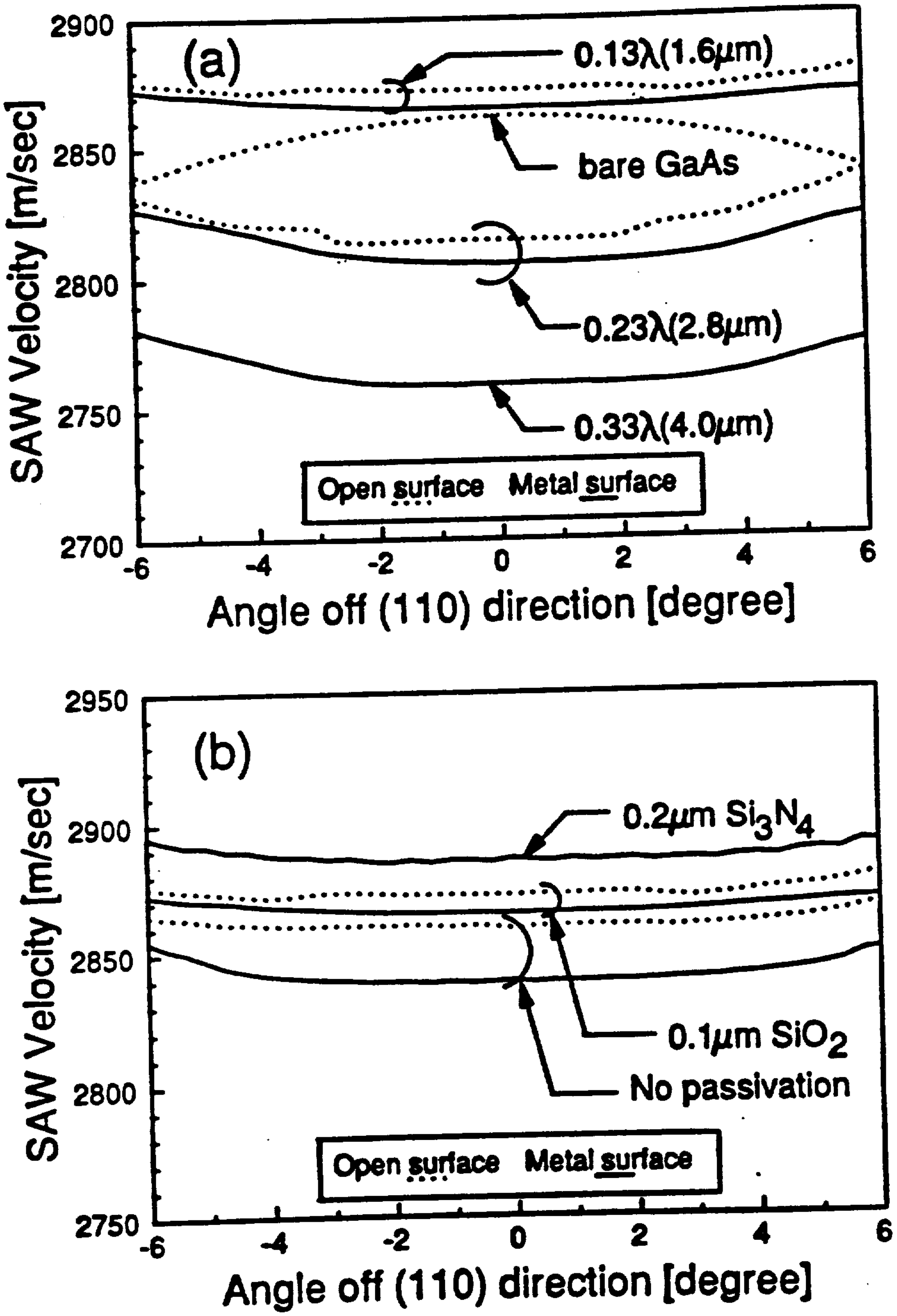

Figure 2.2.6 Experimental velocity surfaces, ZnO films normalized by λ= 12 μm. (a) Effect of ZnO thickness and 0.1 μm SiO_2 layer, (b) Effect of passivation for 0.13λ (1.6 μm) ZnO.

Applying angular spectrum of plane waves (ASPW) theory to the transverse scans produced velocity surfaces shown in Fig. 2.2.6. The usage of the laser probe system for the ASPW theory has been described in detail in Hunt and Hunsinger [2.2.20]. Note that there is close agreement between Figs. 2.2.5 and 2.2.6.

The analysis of the velocity curve in terms of an anisotropy parameter gives an immediate insight into the diffraction of a propagating SAW. Using a parabolic approximation, the velocity of the pure mode direction (<110> of GaAs) can be expressed as

$$v(\theta)=v_o(1 - b\theta^2) \tag{2.2.3}$$

where v_o is the velocity at $\theta=0$, θ is an angle off the pure mode direction in radian, and b is the anisotropy parameter. The beam is spreading for $b<0$ and focussing for $b>0$ comparing to the isotropic case, $b=0$ [2.2.21]. As shown in Fig. 2.2.5, the parabolic approximation for the 0.33λ thick ZnO film is not appropriate since the curvature is almost flat (*i.e.* isotropic) near the zero angle. Therefore, the values of b for the 0.13 and 0.23λ thickness were obtained from a least-square's curve fitting and are listed in Table 2.2.3 for the comparison between the theory and the experiment.

The inversion of the velocity curves was investigated numerically for the other materials. We arbitrarily chose GaAs, GaP, InAs, and InP since they are substrates with cubic symmetry, and AlN and ZnO as the piezoelectric film both of which have hexagonal symmetry. Material constants were available from Auld [2.2.22]. The inversion was observed for all of the combinations; *e.g.*, it occurred at 0.05λ thick ZnO or 0.03λ thick AlN both on GaAs. The curves became flat as the film thickness was increased due to the isotropic characteristic of the films. In addition, SAW modes exist only for relatively thin films of AlN due to its much faster velocity (5608 m/s) than those of the substrates (1990-3540 m/s); *e.g.*, for 0.11λ or less thickness of AlN on GaAs. For comparison, the velocity curve was also calculated for an isotropic film such as fused silica (3411 m/s) on GaAs, but in this case the inversion did not occur. The calculation results suggest that both the hexagonal films (AlN and ZnO) cause the inversion of the velocity surface in a cubic material for the cases mentioned above. This indicates that it is possible to control the diffraction of the substrate with proper selection of film thickness.

2.2.4.2 Line-focus-beam scanning acoustic microscope measurements

Again, Fig. 2.2.5(a) indicates that the velocity surface inversion occurs about the <110> direction when the c-oriented ZnO film is deposited on a {001}-cut GaAs wafer. Our theoretical calculations also reveal that the reverse situation appears if the propagation is along another pure mode axis <100> direction. In order to evaluate this prediction and allow the measurement of acoustic properties along non-piezoelectric directions, velocity measurements using a Line-Focus-Beam scanning acoustic microscope (LFBSAM) 3 operating at 225 MHz were performed. The technique to deduce the velocity is the so-called V(z) analysis which has been reported in detail in [2.2.23]. Due to the requirement of the water coupling between the SAM lens and the sample, the velocity measured is associated with leaky waves for which the waves leak their energy into the water. Thus, the measured velocity due to the water loading is slightly different from the one measured by the laser probe. The mass loading of the water is included in the V(z) analysis. The LFBSAM can also measure both leaky SAW and pseudo SAW [2.2.23] which can exist in a {001}-cut cubic crystal. The advantage of this method is the ability to measure the velocity profile in 360 degrees. However, the spatial

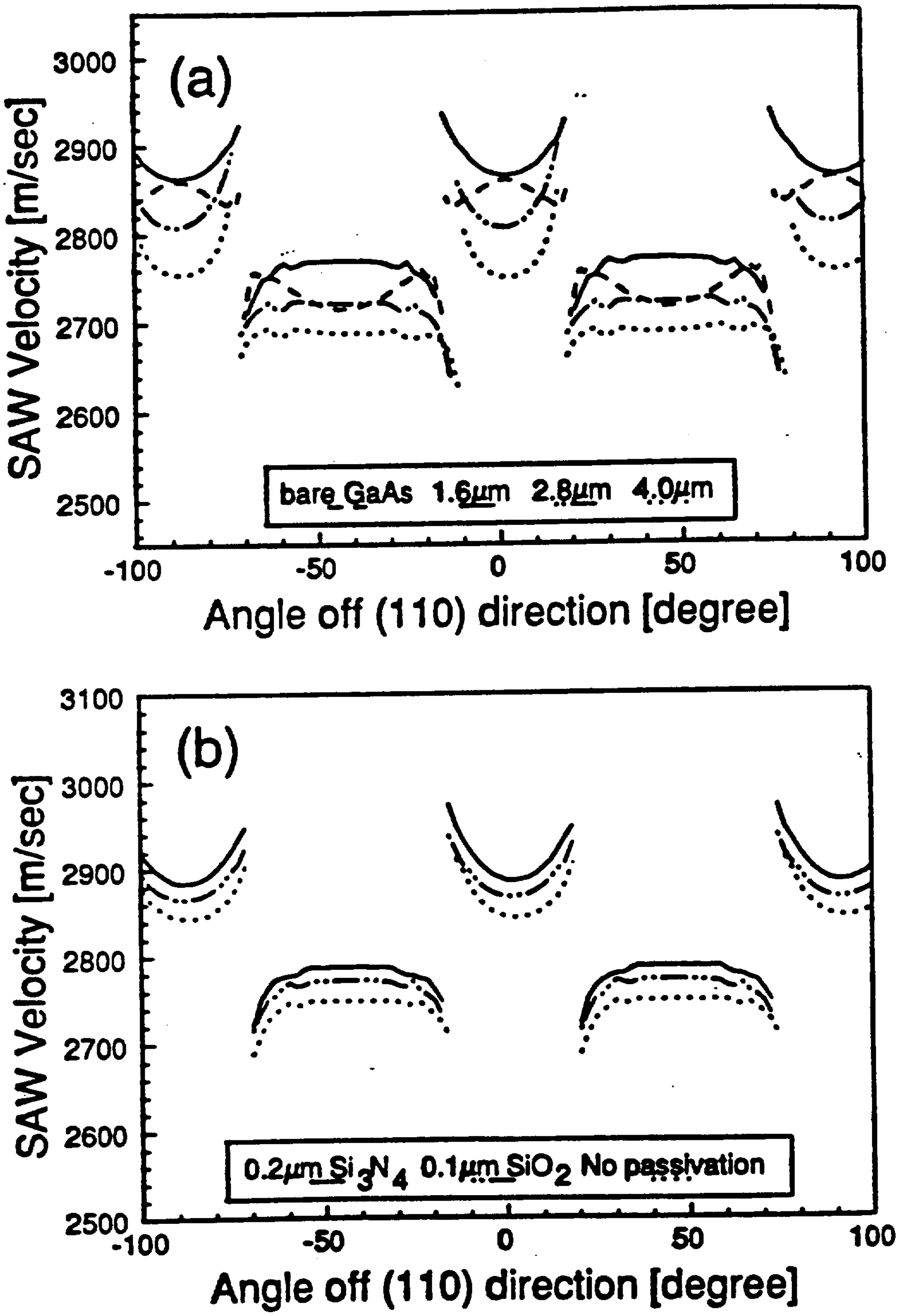

Figure 2.2.7 Velocity surfaces measured by LFBSAM at 225 MHz. (a) Effect of ZnO film thickness with a 0.1 µm SiO_2 layer, (b) Effect of passivation for 1.6µm ZnO film.

resolution is limited (in this case an area of 300 μm by 1000 μm) because of the defocussing requirements [2.2.23]. For the bare {001}-cut GaAs, both leaky SAW and pseudo SAW profiles were reported [2.2.24].

Fig. 2.2.7(a) shows the measured velocity surfaces for the 0 (bare), 1.6, 2.8, and 4.0 μm ZnO film on {001}-cut GaAs. It is seen that the ZnO films change the focussing behavior about the <110> direction into the non-focussing one and the reverse inversions appear about the <100> direction. These results support our theoretical results and our laser probe measurements. Effects of the passivation layer (0.1 μm SiO_2 or 0.2 μm Si_3N_4) for 1.6 μm ZnO and {001}-cut GaAs on the velocity surface are provided in Fig. 2.2.7(b), and they also agree with the results obtained using the laser probe. Note that Figs. 2.2.7(a)(b) are solely used to demonstrate the leaky SAW behavior about the <110> and <100> directions, the details of their behavior farther away from these angles are not shown.

Experimental velocity surfaces with and without the metallization pad for small deviations about the <110> direction are given in Fig. 2.2.8. Because the velocity measurement accuracy performed by LFBSAM is better than ±0.02%, in principle these data can be used to deduce the value of K^2. Recently, the V(z) analysis of LFBSAM measurements for a multilayered anisotropic structure has been reported [2.2.25]. However, due to a lack of understanding of the dielectric loading of the piezoelectric effect by the water, measurements of K^2 can not be carried out at present. On the other hand, the water is distilled and de-ionized to have a resistivity of $10^7\Omega$-cm. The conductivity effect of the water is mingled with its dielectric effect because the characteristic electromagnetic impedance is equal to the root square of the permeability over the permittivity.

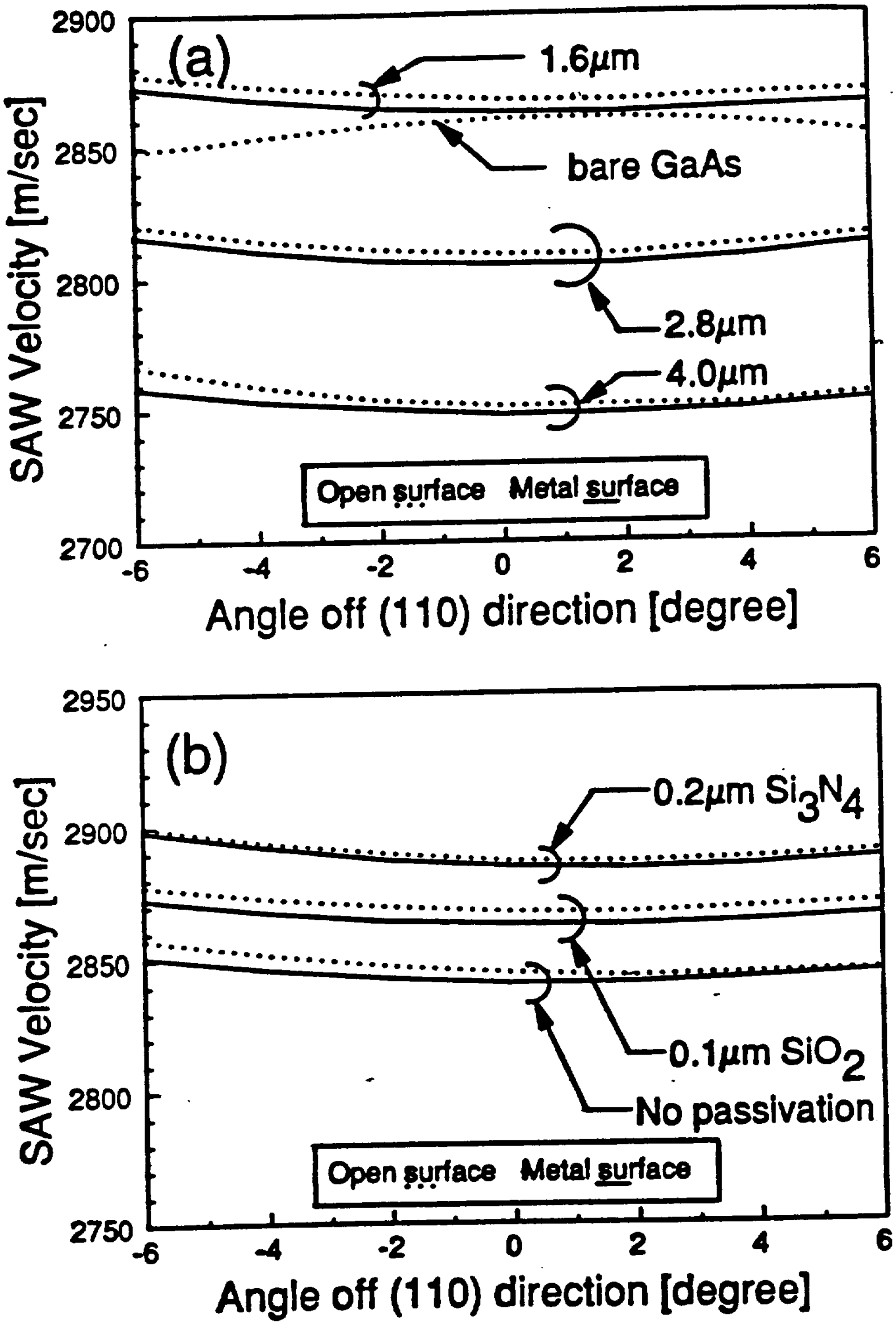

Figure 2.2.8 Velocity surfaces measured by LFBSAM for small deviations about <110>. (a) Effect of ZnO thickness with 0.1 µm SiO_2 layer, (b) Effect of passivation for 1.6µm ZnO film.

2.2.5 Reflection properties of metallic gratings

An exact theoretical analysis of the reflection properties is extremely difficult. The reflectivity and velocity shift of metal gratings on a single crystal can be analyzed by using Datta and Hunsinger's technique [2.2.3] if the metal thickness is small compared to the SAW wavelength. Considering that the metallic gratings effect the surface, Datta and Hunsinger's theory developed for a single material can be directly applied to a multilayered substrate by obtaining effective material constants of a single material equivalent to a multilayered substrate. The effective constants c_{eff} may be obtained by integrating multilayered material constants weighted with SAW energy distribution and normalized by the energy as follows,

$$c_{eff} = \frac{\int dy\; c(y) P(y)}{\int dy\; P(y)} \tag{2.2.4}$$

where c(y) is a constant of the multilayered material tensor and P(y) is the energy distribution as a function of the depth y. Since the substrate is piezoelectric, the mechanical, electro-mechanical, and electrical energy distribution are utilized for stiffness, piezoelectric, and permittivity tensor, respectively. The detailed expression for the energy distribution is available from the Laguerre polynomial technique [2.2.7] and is not reproduced here. The validity of using the effective constants may be checked by comparing between the SAW velocity of ZnO/GaAs and that of the equivalent single material as is shown in Fig. 2.2.9. The maximum difference between both velocities is 0.009.

The reflectivity of gratings can be divided into two components: the piezoelectric and the mechanical. The former depends on the piezoelectric coupling constant K^2, and the latter depends on the thickness h of the metal in wavelength λ. The reflectivity r of a metal strip

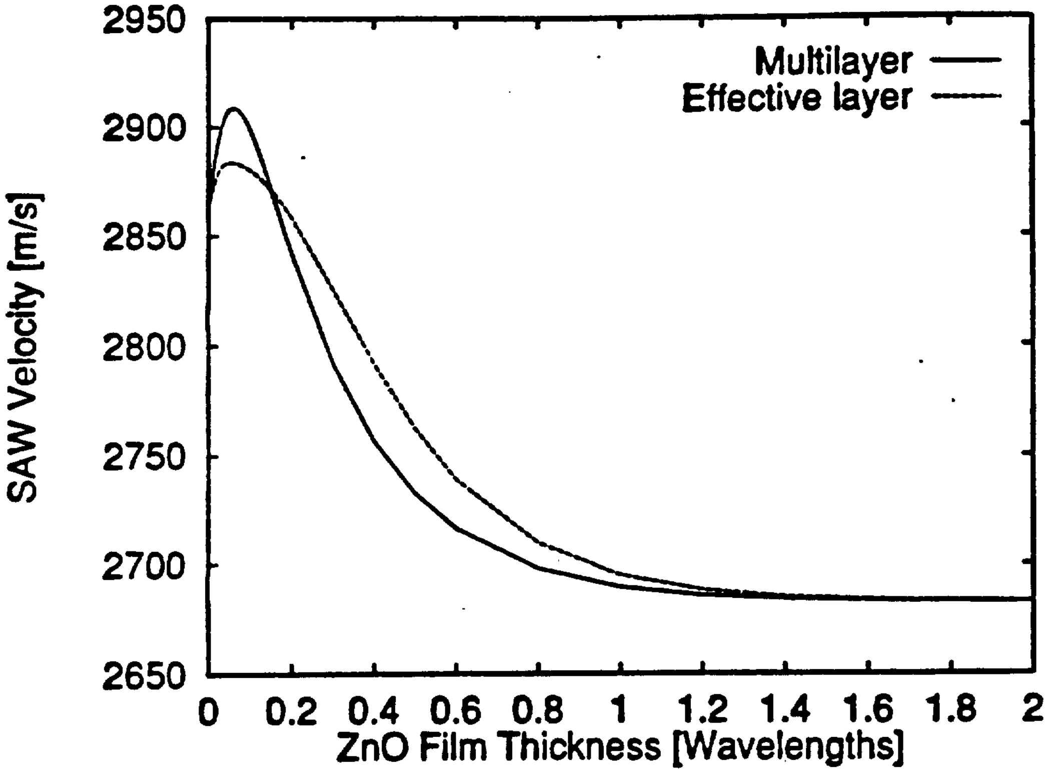

Figure 2.2.9 A comparison between SAW velocity of multilayer (ZnO/GaAs) and that of effective single material.

with a metallization ratio η at a frequency f can be given by

$$r = j\left[P_z(\eta)\frac{K^2}{2} + F_z\frac{h}{\lambda}\right]\sin\frac{\eta\pi f}{f_o} \qquad (2.2.5)$$

where f_o is the center frequency, P_z is the piezoelectric scattering coefficient, and F_z is the first order mechanical scattering coefficient. P_z is the same for all substrate-electrode combinations and depends only on the metallization ratio. On the other hand, F_z depends only on the substrate and electrode materials and is independent of the metallization ratio. More detailed explanation about these parameters can be found in Hunsinger [2.2.3].

The value of P_z at f_o is given as a function of η by using the Legendre function [2.2.3]. It is to be -0.75 when the value of η is 0.5. Fig. 2.2.10 shows the calculated values of K^2 and F_z for an aluminum grating with 1/4 λ width and 1/4 λ spacing (the same geometry for solid finger interdigital transducers (IDTs)) on ZnO/0.25 μm Si_3N_4/GaAs as a function of ZnO film thickness. As shown in Fig. 2.2.10, the value of K^2 increases from 0.07 % (GaAs) to 1.0 % (ZnO), and that of F_z decreases from 0.9 (GaAs) to 0.2 (ZnO) as the film thickness increases. Thus, it is possible for the piezoelectric scattering and the mechanical scattering to cancel each other due to their opposite signs. This zero reflectivity can not only reduce triple transit interference (TTI) of IDTs for some filter applications but also allows the use of a single finger IDT instead of a split finger IDT, thus increasing the frequency range of easily fabricated devices.

In order to verify this remarkable theoretical expectation, a set of characterization devices shown in Fig. 2.2.11 was designed to measure the effect of the gratings. The ZnO films were grown by dc sputtering method with thicknesses of 0.8 and 1.6 μm with a 0.25

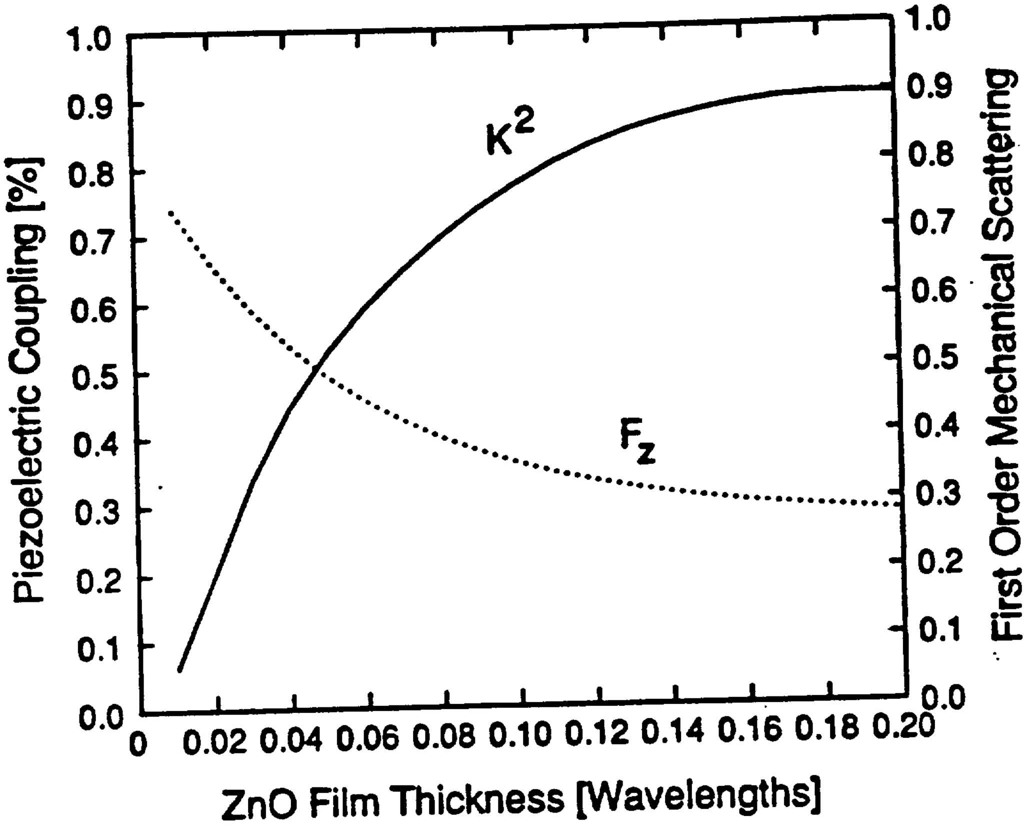

Figure 2.2.10 Piezoelectric coupling K^2 and first order mechanical scattering coefficient F_z of aluminum gratings on ZnO/0.25 µm Si_3N_4/GaAs.

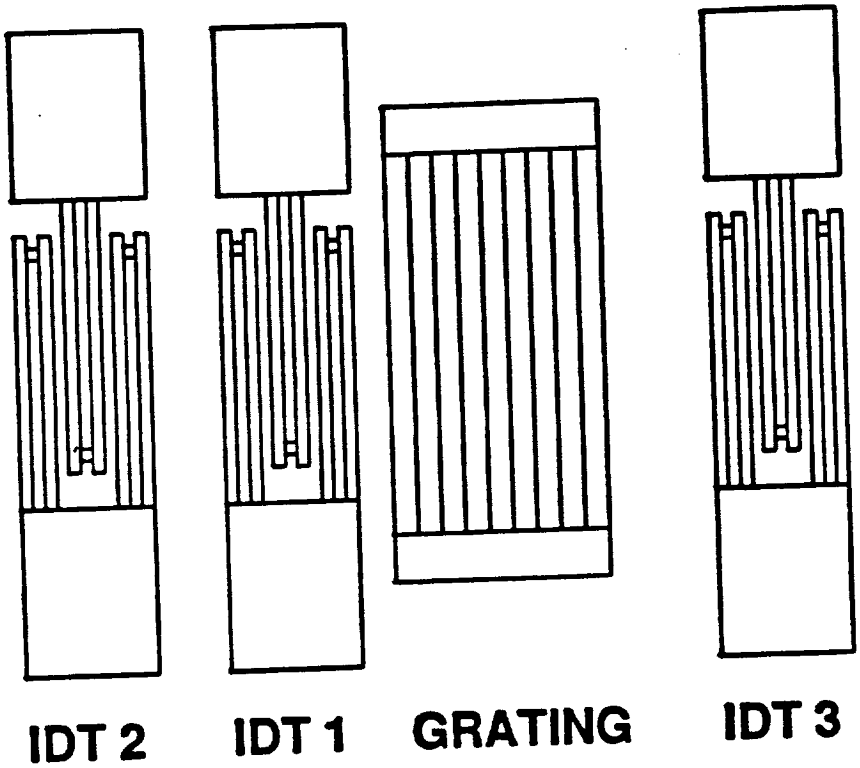
IDT 2
IDT 1
GRATING
IDT 3

Figure 2.2.11 Device configuration for reflectivity measurements.

μm CVD grown Si_3N_4 passivation layer. The probe IDTs are 100 λ wide where λ is 16 μm with a center frequency around 180 MHz. Each of the probe IDTs is 20 λ long, and split finger electrodes are employed in order to reduce internal reflections. Both shorted and open gratings with the width of 108 λ are investigated, and each of the gratings consists of 200 strips with 1/4 λ width and 1/4 λ spacing. The distance from the center of the front two IDT electrodes with same polarity to the edge of the shorted grating is designed to be (1/2n+1/4) λ and 1/2nλ, where n is an integer, to measure the sign of the reflectivity since the phase construction of a standing wave occurs at the distance of (1/2n+1/4) λ for a negative sign of reflectivity and $1/2n \lambda$ for a positive one [2.2.26].

The properties of open gratings were also investigated along with the shorted gratings. Unlike the shorted gratings, the SAW regeneration from the open grating should be considered in the analysis. At the center frequency, the reflectivity, r_o, of an open metal strip with the metallization ratio of 0.5 is given as [2.2.27]

$$r_o = r + j0.64K^2 \tag{2.2.6}$$

Due to this regeneration, the reflectivity of the open gratings always retains a positive value with a larger magnitude than that of the shorted gratings; thus, the open gratings may be suitable for reflectors or resonators, for which a strong reflectivity is desirable.

The identical IDT probes with free and metalized surface with a 120 λ wide and 200 λ long metal pad instead of the gratings were also fabricated to calibrate the measured data. Moreover, from these calibration sets, the SAW properties of velocity, K^2, and attenuation for the ZnO/GaAs substrates were also characterized. Three different thicknesses of the aluminum gratings were deposited with an e-beam evaporator. The measurements using a surface profile measurement apparatus showed they were 400, 800, and 1200 Å, which corresponded to 0.25, 0.50, and 0.75 % of λ, respectively. The measurements were performed with a HP8753C network analyzer with a time-gating feature.

The velocities for the 0.8 μm and 1.6 μm ZnO thicknesses were 2928 m/sec and 2912 m/sec, respectively, from the center frequency measurements. The value of K^2 was also obtained by comparing the difference of the time delay from the IDT 1 and 2 to the IDT 1 and 3 between the free surface and the metalized surface. It was 0.73 % and 1.1 %, respectively, but their accuracies may be degraded by the inaccurate value for the free surface velocity obtained from the center frequency measurements due to the velocity shift under the IDTs. More accurate measurement using the laser probe or the LFBSAM was not attempted for these substrates, but the measurements for the other substrates are already discussed in the previous section. The propagation loss was also measured by comparing insertion loss of the IDT 1 and 2 pair and that of the IDT 1 and 3 pair with a similar manner described in Melloch and Wagers [2.2.18]. Note that the loss due to diffraction is not considered in this technique. The loss of the 0.8 μm and 1.6 μm thickness was obtained to be 0.7 dB/μs and 0.8 dB/μs on the metal pad, and 2.6 dB/μs and 3.6 dB/μs on the free surface, respectively. This difference between the free surface and the metal pad is due to the large diffraction of the ZnO/GaAs. The 120 λ wide metal pad must act as a waveguide to confine the acoustic energy and reduce the insertion loss, resulting in the reduction of the propagation loss. Note that the results coincide well with the measurements by the laser probe, which are discussed in the previous section.

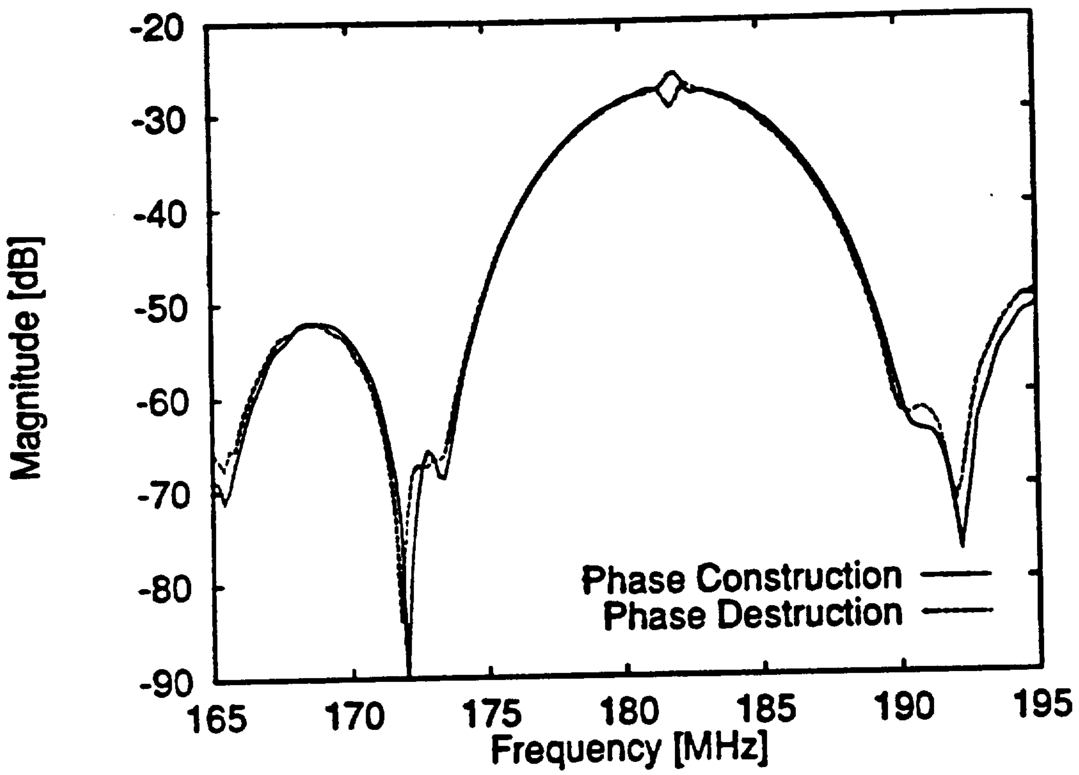

Figure 2.2.12 Typical frequency response of two shorted gratings with the same metal thickness showing phase construction and destruction.

Fig. 2.2.12 shows typical unmatched frequency response between the IDT 1 and 2 pair for the two shorted gratings (as mentioned before, the difference of their distance from the IDT 1 is 1/4 λ) with the same metal thickness. Note that the phase construction and destruction are clearly shown about f_o. The reflectivity of the grating was obtained using the time gating technique described in Wright [2.2.27]. Comparing the response of the IDT 1 and 2 pair with the grating to that without gives a total reflectivity of the gratings assuming the distance between the IDT 1 and the grating is exactly correct for the total phase construction and destruction. Note that the propagation loss is negligible. The reflectivity of each electrode of the grating is given by [2.2.26]

$$|r| = \frac{\tanh^{-1}(R)}{N} \qquad (2.2.7)$$

where R is the total reflectivity and N is the number of the electrode.

Figs. 2.2.13 (a) and (b) show the measured data and theoretical calculation for the reflectivity of an electrode in the shorted and the open grating at f_o as a function of the metal thickness normalized in λ. Prior work concerning gratings on GaAs can be found in the literature[2.2.28]. The value of K^2 was assumed to be 70 % of the theoretical value based upon the measurements presented in the previous section. From the theoretical data, it is seen that for the shorted gratings the positive reflectivity on bare GaAs decreases, and eventually it becomes negative. The change in the reflectivity of the open gratings with film thickness is not so drastic as for the shorted gratings. On the other hand, the reflectivity on bare GaAs at f_o is given by 1.1 h/λ, and its value has a marginal difference between the shorted and the open gratings due to the very weakly piezoelectric regeneration. From the measurements, it

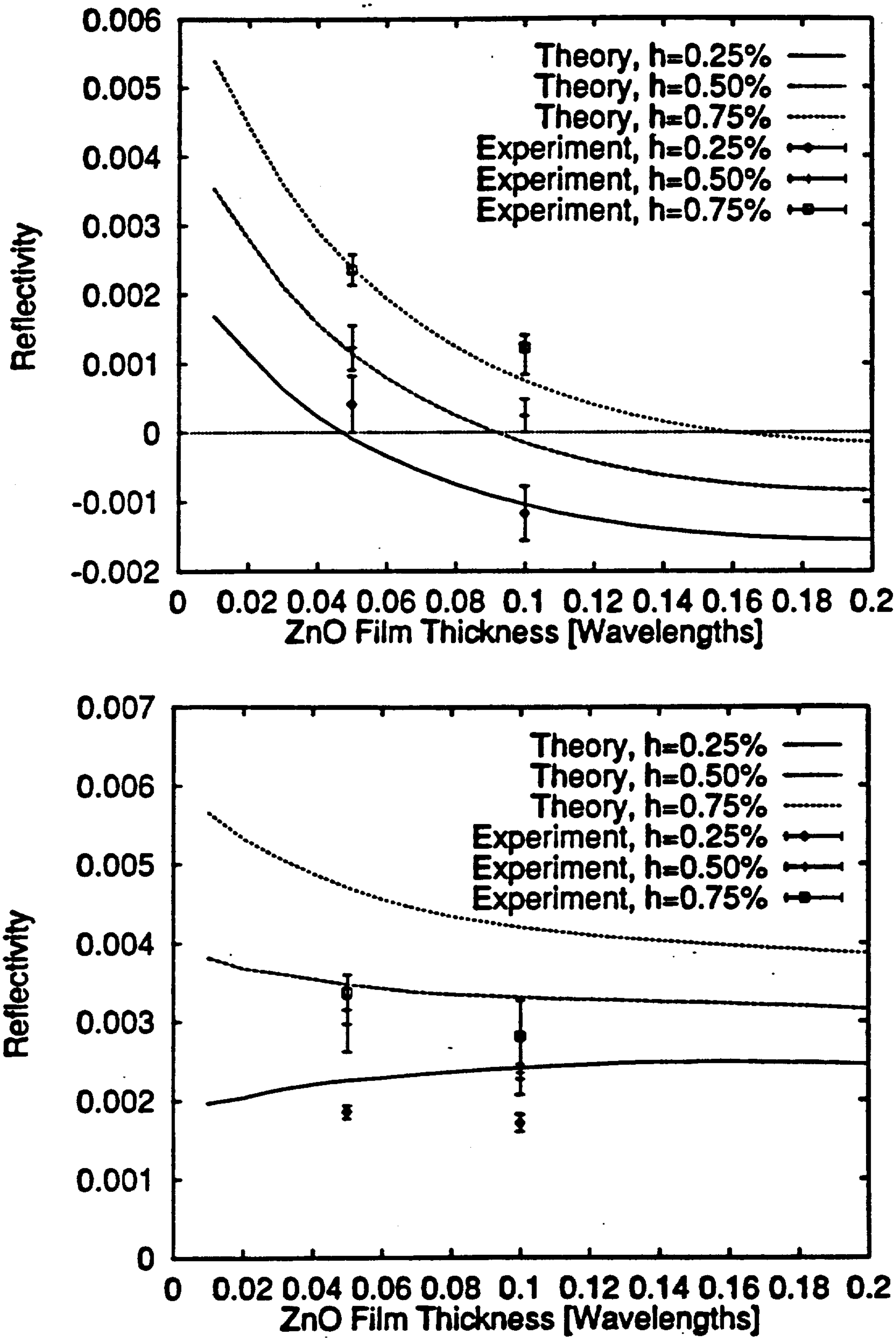

Figure 2.2.13 Reflectivity of an aluminum electrode with 1/4λ width and spacing 1/4λ at f_o where the metal thickness in λ is represented by h: (a) shorted grating, (b) open grating.

should be noted that the sign of the reflectivity changes depending upon the metal thickness and zero reflectivity occurs at the metal thickness of 0.50 % λ for 0.1 λ (1.6 μm) thickness of the ZnO film for the shorted grating while the reflectivity for the open grating is always positive. The measured data shows relatively good agreement with the theoretical prediction. The difference between the theory and the measurements may be due to the inaccurate effective material constants as well as the variation in the uniformity of the film and the metal thickness over the entire surface.

The measurement of velocity shift of the grating was attempted from the shift of the center frequency of the grating only response, f_{og}. The shift of f_{og} depending upon the metal thickness was not noticeable for both ZnO films, but the frequency shift between the shorted and the open grating was observed with the value of 0.05-0.25 MHz. It is known that the regeneration of SAW from the open grating causes an increase in the velocity by 0.25 K^2 at $\eta=0.5$ [2.2.27]; thus, the measured values are comparable to the calculated ones of 0.18 MHz and 0.25 MHz for 0.8 μm and 1.6 μm thickness of the ZnO film, respectively.

2.2.6 Summary of Thin Film Work

Theoretical and experimental data have been provided for the design of ZnO/GaAs based SAW devices, which may have the potential application for the monolithic integration of SAW and electronic devices. An analytical tool for SAW properties of a piezoelectric thin film was developed in this study, and its accuracy was verified with the experimental results of ZnO/GaAs substrates.

The SAW properties of both rf magnetron and dc triode sputtered ZnO films on {001}-cut <110>-propagating GaAs have been measured and compared to theoretical calculations. The passivation layer improves the ZnO film quality appreciably for both sputtering methods by reducing the propagation loss 0.5-1.3 dB/μs at 240 MHz depending upon the ZnO film thickness. The best quality of film observed has been 1.6 μm dc triode sputtered ZnO film with 2000Å Si_3N_4, for which the propagation loss was 0.9-1.0 dB/μs at 240-280 MHz. Its value of K^2 would be larger than that of the film without the passivation layer, 0.65-0.75%. Compared with bare GaAs, this particular film substrate has approximately 0.6 % lower SAW velocity, at least 10 times larger K^2, and 0.6 dB/μs more propagation loss.

The velocity surfaces of the ZnO/GaAs substrates have been measured using a knife-edge laser probe and a LFBSAM and have been theoretically predicted. It is found that for certain thicknesses of ZnO film the velocity surfaces about the <110> direction are concave upward in contrast with the concave downward curvature of the bare GaAs. This causes the SAW to diffract as soon as it leaves the transducer. Theoretically, c-axis oriented hexagonal films such as AlN and ZnO can be used to control the diffraction of a cubic substrate such as GaAs, GaP, InAs, and InP by varying the film thickness.

The reflectivity of shorted and open metal gratings has been analyzed and measured. The analysis showed that zero reflectivity could be obtained by a combination of the ZnO film and the metal thickness and the metallization ratio. Experimentally, zero reflectivity at the center frequency was observed at 0.5 % λ thickness of a shorted aluminum grating with 1/4 λ width and 1/4 λ spacing on 1.6 μm thick ZnO film with 0.25 μm thick Si_3N_4 passivation layer over GaAs.

TABLE 2.2.1: ZnO Film Deposition Parameters

Sputtering method	**RF magnetron**	**DC Triode**
Target	ceramic 6.5 inch dia.	ceramic 4 inch dia.
Target to substrate	1.5 inches	2.5 inches
Background pressure	8 mTorr	3 mTorr
Substrate temperature	350° C	250° C
Gas	82% Ar 18% O_2	90% Ar 10% O_2
Power	400W	150W
Rate	4.6 μm/hour	1.4 μm/hour

TABLE 2.2.2: Number of Finger Pairs N is 50.5 for all IDTs

No. i	**Type**	**Nominal wavelength (λ_i) [μm]**	**Measured center freq.(f_c) [MHz]**	**Aperture (W_i) [μm]**	**Aperture (W_i) [λ_i]**
1	split	16	175.0-182.2	80	5
2	solid	12	230.2-240.6	100	8.3
3	solid	10	273.5-287.4	100	10
4	solid	8	339.0-356.8	80	10

TABLE 2.2.3: Anisotropic Factor of 0.13 and 0.23λ Thickness of ZnO Film. The Values Are Obtained by Curve Fitting With a Least-Square Method. The Values in the Parentheses Are for the Short Surface.

Zno Thickness	**Sputtering method**	**Passivation**	**Theory**	**Experiment**
0.0	----	0	0.22	0.45[a]
0.13λ	RF Magnetron	1	-0.13 (-0.16)	-0.18 (-0.17)
0.13λ	DC triode	2	-0.13 (-0.16)	N/A (-0.19)
0.13λ	DC triode	0	-0.14 (-0.17)	-0.22 (-0.4)
0.23λ	RF Magnetron	1	-0.44 (-0.56)	-0.71 (-0.55)

0: None, 1: 0.1μm SiO_2 , 2: 0.2μm Si_3N_4

[a] Reference [2.2.11]

2.3 Acoustoelectric Amplifiers

In the previous report, the concept of acoustoelectric amplification was introduced, an analytical model was briefly presented and the motivation for the research in this area was developed. In this section we present some theoretical predictions for bilevel structures using the model of Henaff et. al [2.3.1]. We also discuss the segmented structure which will be fabricated and tested over the first quarter of 1995. Finally the application of these devices to the HACT imager is discussed including size and power consumption considerations.

2.3.1 Bilevel Model

The simple model of [2.3.1] was adopted from bulk acoustic (BAW) wave amplifiers [2.3.2] and extended with great success to surface acoustics wave (SAW) devices. Due to its relative simplicity and accuracy this model found widespread use in modelling devices consisting of a semiconductor layer in close proximity or deposited on a piezoelectric substrate. It is also valid for a piezoelectric film deposited on a semiconductor or a bulk piezoelectric semiconductor. The details of the derivation may be found in [2.3.1].

Figure 2.3.1 shows schematically a amplifier structure on bulk n-type GaAs. The wave is launched and received by the interdigital transducers (IDTs) at the ends of the device and the surface wave is amplified due to the interaction with the drifting carriers accelerated by the voltage applied at the ohmic contacts. Figure 2.3.2 shows a comparison between the bilevel model and the experimental measurements on a 10^{13} O_2 doped GaAs substrate at a synchronous frequency of 176 MHz [2.3.3]. Data points were taken from Figure 3 of [2.3.3] and smoothed using the SPLINE function in MATLAB to produce a curve for the comparison

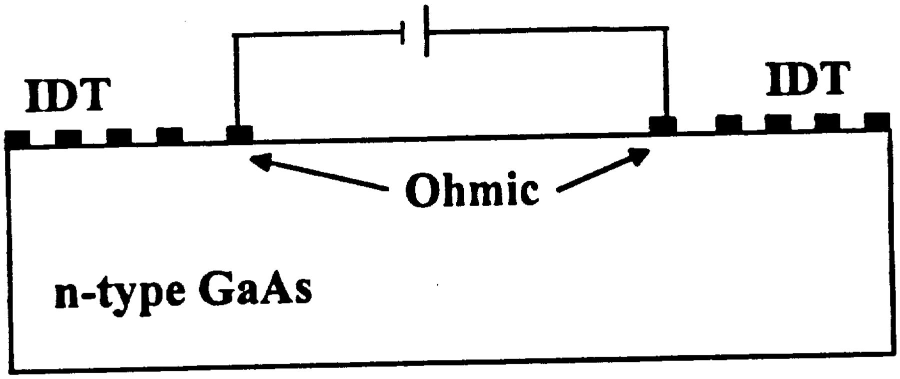
IDT
IDT
Ohmic
n-type GaAs

Figure 2.3.1 Schematic diagram of the amplifier structure on bulk GaAs.

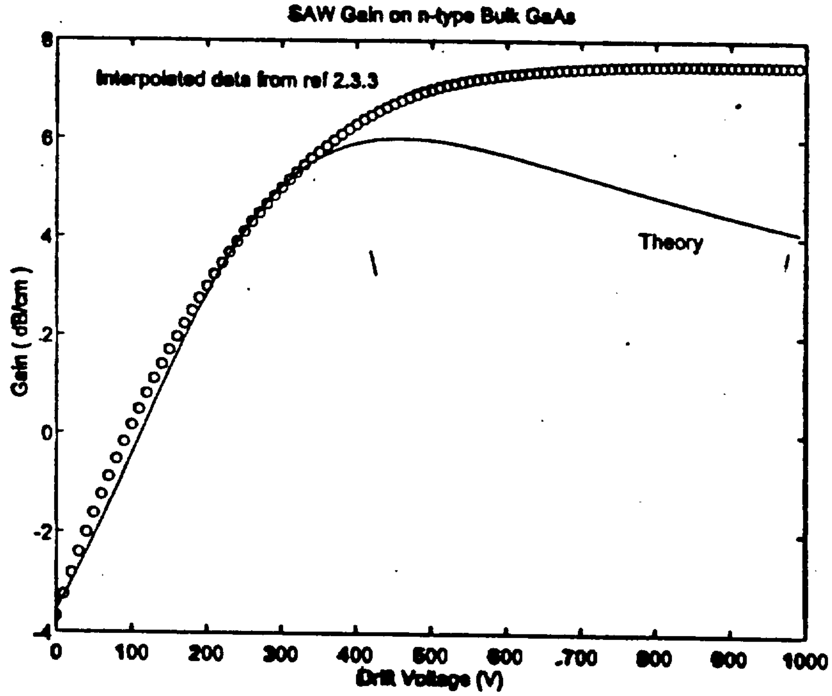

Figure 2.3.2 Comparison of the bilevel model and experimental measurements of GaAs substrate.

with the bilevel model. The y axis is the gain measured over a 1 cm separation of the ohmic contacts and the x axis is the voltage applied between the contacts. The model gives a good first order approximation to the experimental device and serves as a good tool to make predictions for our first experimental designs. The model deviates from the experiment at high voltage but more accurate models will be developed as required throughout the project.

2.3.2 Segmented Amplifier

The device of Figure 2.2.1 had several undesirable characteristics, the most prominent being the large voltage required across the ohmic contacts to amplify the acoustic signal.. This high voltage problem was resolved by using a segmented structure [2.3.4]. The amplifier is segmented into many smaller sections over which a much smaller voltage is applied. The length of the segment is defined as L1 and the segment separation as L2 as illustrated in Figure 2.3.3. The gain of the structure is the difference between the sum of the gain in the L1 sections and the loss in the L2 sections. Such structures were demonstrated successfully on CdS with L1= 500 um and L2 = 250 um [2.3.5]. The voltage for maximum gain for these devices occurred around 120 V.

Our approach is to design devices with the shortest possible segment length to give the minimum power dissipation. Devices have been designed with L1 on the order one acoustic wavelength to enable the device to be operated with drift voltages less than 5 V. The bilevel model has been extended to predict the gain of the segmented structure. As a comparison to the high voltage device, the curve of Figure 2.3.4 shows the predicted performance of a segmented structure on 10^{13} doped bulk GaAs with L1 =100 um and L2 = 20 um at a

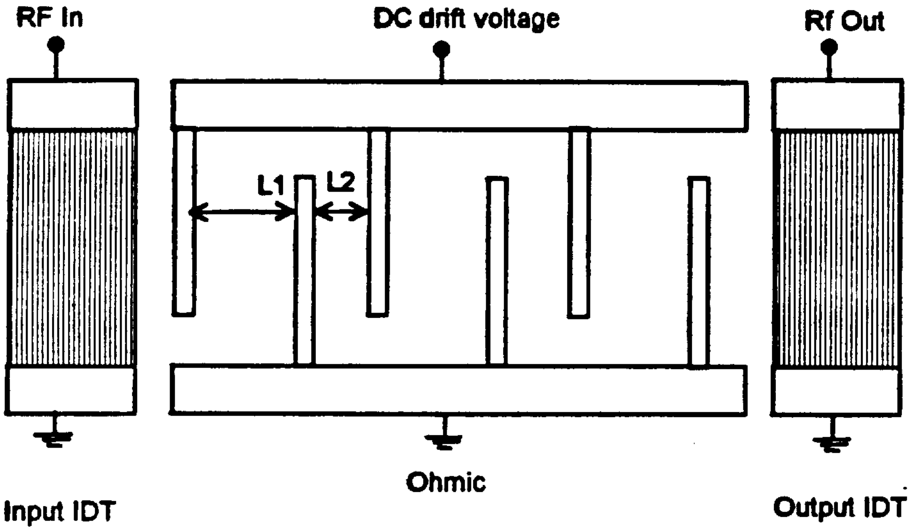
RF In
DC drift voltage
Rf Out
L1
L2
Input IDT
Ohmic
Output IDT

Figure 2.3.3 Proposed amplifier design using smaller segments.

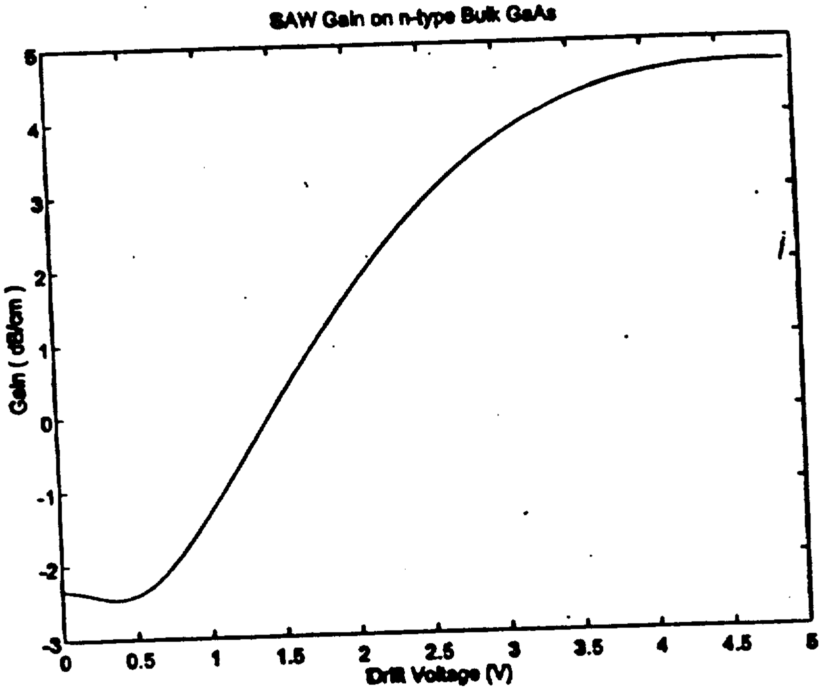

Figure 2.3.4 Bilevel model predictions of the performance of the segmented amplifier design.

synchronous frequency of 176 MHz. In order to achieve low voltage operation, some tradeoffs in size and power consumption must be made. As shown in Figure 2.3.4, the gain per unit length is smaller for the segmented structure than the high voltage device due to the loss incurred in the L2 sections. The power consumption may also be higher depending on the length of L2. These tradeoffs will be examined in more detail as the project progresses.

Our devices will be fabricated on substrates consisting of a n-type epilayer on top of intrinsic GaAs. The device operation is the same as for bulk material and the bilevel model has facilitated the design of the first experimental pass. The layer stackup will be consistent with that used in the HACT imager so that the amplifier may be integrated into the imager as will be discussed in a later section. The structures intended for use in the HACT will consist of single L1 gain sections placed throughout the imager. The segmented structure serves as an experimental device to develop an accurate model and is also intended for signal processing applications as a stand-alone device or integrated into a IC with other functions.

2.3.3 ZnO Overlay

The advantage of using a GaAs substrate for this device is the high carrier mobility. In the region where the carrier velocity is less than the saturation velocity, the carrier velocity is equal to the product of the mobility and the applied electric field. The gain of the device is proportional to the carrier velocity and so is proportional to the applied electric field (ie. applied voltage) times the mobility. Out of this it may be seen that the higher the mobility the lower the applied required to obtain the same gain.

The gain is also proportional to the electromechanical coupling coefficient, k^2. The disadvantage to using GaAs is the relatively low value of k^2 compared to substrates such as $LiNbO_3$. However this problem may be overcome by using a ZnO overlay to increase the k^2 by at least an order of magnitude. Figure 2.3.5 shows the gain vs voltage curve for a ZnO overlay with an estimated $k^2 = 0.5\%$ on the device modelled in Figure 2.3.4. We observe a peak gain improvement of approximately 35 dB with the addition of the ZnO film.

2.3.4 Application to HACT Imager

As mentioned in the previous report, the motivation behind this project is to enable the insertion of small gain blocks in between the pixels of the HACT imager to maintain a constant SAW amplitude across all of the pixels. Our plan is to use minimum length blocks to maintain minimum voltage operation. For example, if we use a single 8 um long gain block at 360 MHz (1 wavelength) with the ZnO overlay the peak gain would be 0.07 dB and the power dissipation would be 0.036 mW. This by itself is not sufficient to overcome the SAW propagation the loss over the length of the imager. However, if for example we insert 100 of these miniature blocks along the propagation path of the SAW, then we would have 7 dB of gain at a power dissipation of 3.7 mW. The number of blocks will be optimized depending on the loss incurred over the SAW propagation path and the available real estate on the imager.

2.3.5 Summary

The bilevel model has been demonstrated for a bulk GaAs substrate. The model is equally applicable to multilayer piezoelectric structures. A segmented structure must be used

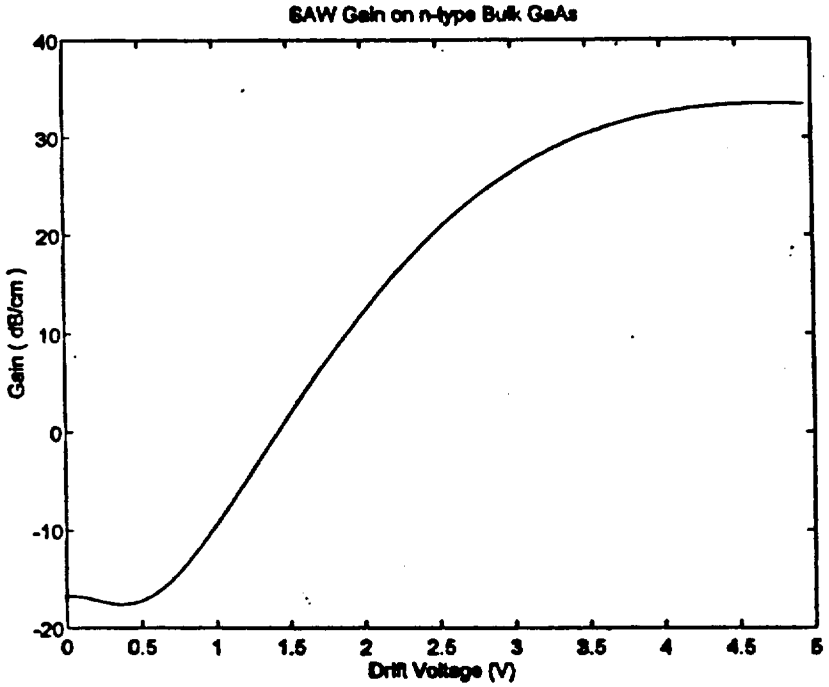

Figure 2.3.5 Gain versus voltage for ZnO overlay on GaAs.

in order to use this device with a practical voltage. Some gain and power consumption tradeoffs which result from the segmented structure will be examined later. The small gain sections will provide a method by which the SAW amplitude may be maintained at a constant level across the length of the HACT imager resulting in an overall lower power consumption for the device.

2.4 Advances in ACT Device Modelling

Two different approaches to the modelling of acoustic charge transport devices have been accomplished. The first is the optimization of the heterojunction acoustic charge transport device. Material layer thicknesses and doping concentrations were varied in an attempt to provide optimal characteristics. The second is modification of the existing code to simulate ZnO on silicon acoustic charge transport devices.

2.4.1 GaAs Heterojunction ACT Optimization

Previous reports detailed the computer model used for the simulation of acoustic charge transport devices and the design goals required for successful operation. One such device structure is displayed in Figure 2.4.1. These computer simulations of the layer structure displayed high charge carrying capacity. However, it was determined that due to the thin AlGaAs layer and the relatively deep location of the charge packet that non-destructive sensing contacts on the surface of the device would be highly inefficient. The image charge created by the signal electrons would appear in the substrate, as shown in Figure 2.4.2, instead of the metal layer. This is due to the distance from the electron packet surface being three times larger than the distance from the packet to the substrate electrons. In order to account for this short coming in the previous design a new design with a thick layer of AlGaAs was examined, Figure 2.4.3.

Simulations of the new structure revealed the surface charge insensitivity of the previous design with essentially the same charge carrying capacity.

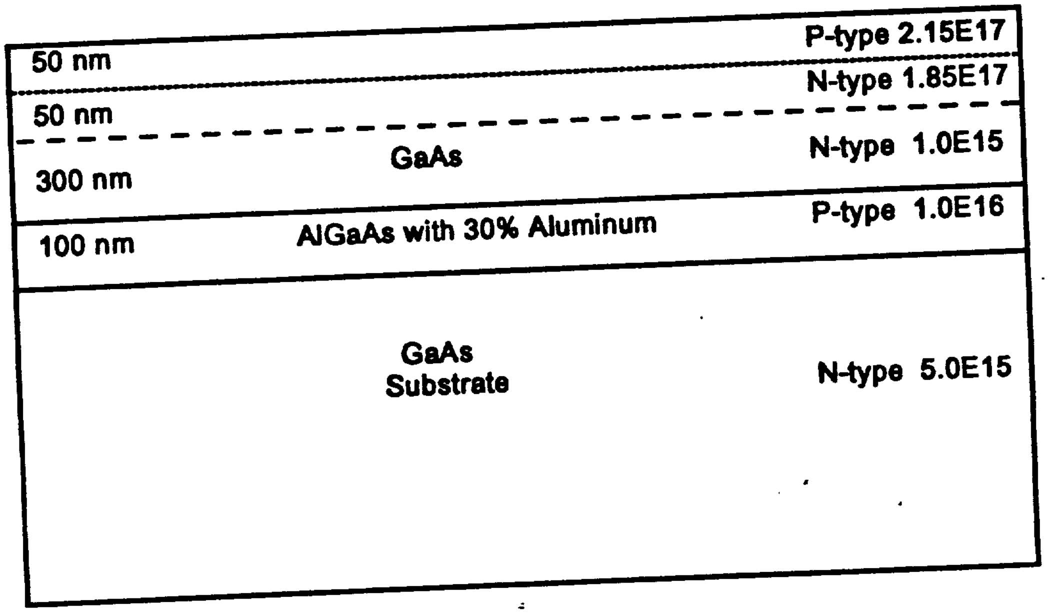
50 nm
P-type 2.15E17
N-type 1.85E17
50 nm
GaAs
N-type 1.0E15
300 nm
100 nm
AlGaAs with 30% Aluminum
P-type 1.0E16
GaAs
Substrate
N-type 5.0E15

Figure 2.4.1 Material layer structure of a HACT device insensitive to fluctuations in the surface state density.

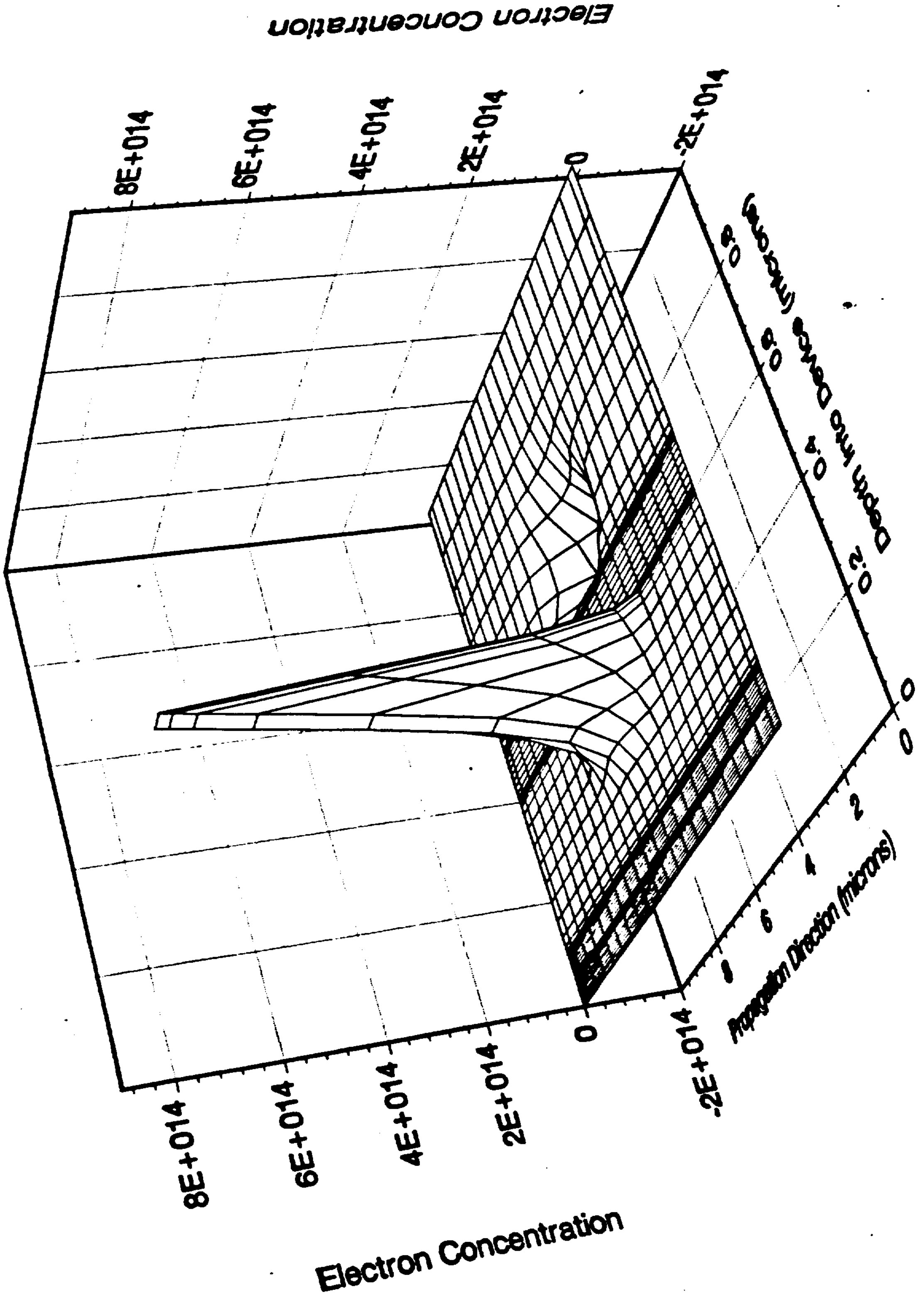

Figure 2.4.2 Signal charge (electrons) in the channel and the associated image charge created in the GaAs substrate.

C-2

10

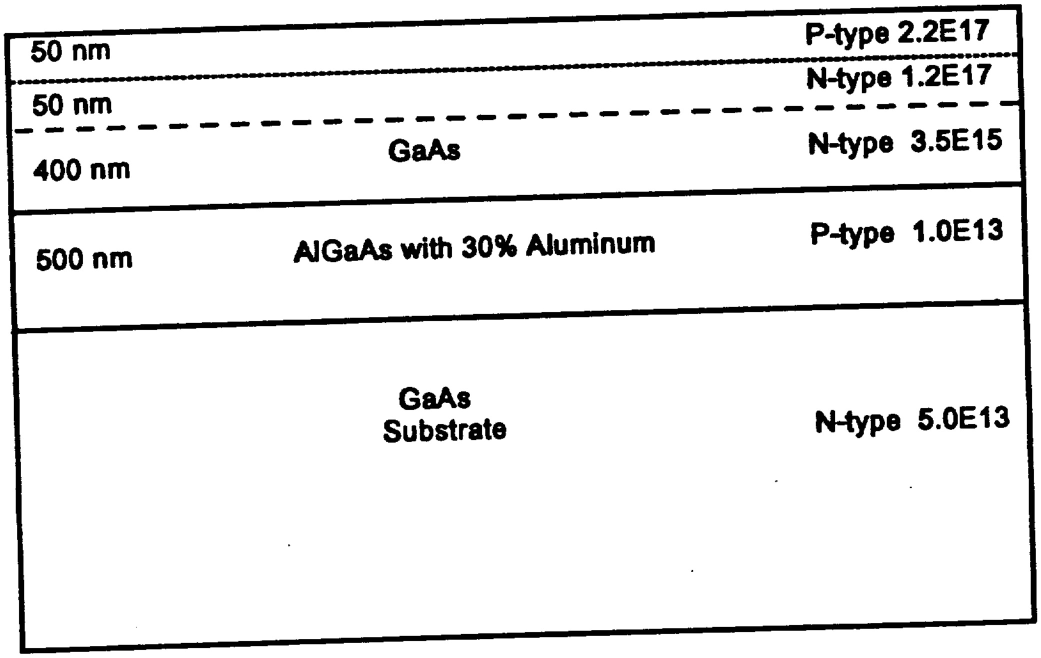
50 nm
P-type 2.2E17
50 nm
N-type 1.2E17
400 nm
GaAs
N-type 3.5E15
500 nm
AlGaAs with 30% Aluminum
P-type 1.0E13
GaAs
Substrate
N-type 5.0E13

Figure 2.4.3 New material layer structure for a HACT device which is insensitive to the surface state concentration and will support image charge at the surface due to the thick AlGaAs layer.

2.4.2 ZnO on Silicon Acoustic Charge Transport

Modifications of the device simulation code have enabled simulations of ZnO thin layers on silicon substrates. Since silicon is not piezoelectric, the acoustic wave is supported in the ZnO layer and the induced potential extends into the underlying silicon. However, since ZnO is highly resistive simulation of the layer within the device model would be very difficult due to the low carrier concentration. The modification of the code enables the simulation of the acoustic wave only in the ZnO, and both the acoustic potential and the carrier concentrations in the silicon substrate. Figure 2.4.4 displays the material structure, and Figure 2.3.5 shows a conduction band contour plot created in the silicon by the SAW potential. The SAW has created a potential well, very similar to that previously reported in GaAs, which is capable of transporting charge.

2.4.3 Future Work on Device Simulation

Future work in the area of acoustic charge simulation will require further optimization of the ZnO/Si system devices. In addition, future code modifications should make it possible to simulate moving charge packets by using a moving simulation domain solution. In this manner synchronis potential effects and injection phenomena maybe modelled leading to highly efficient device designs.

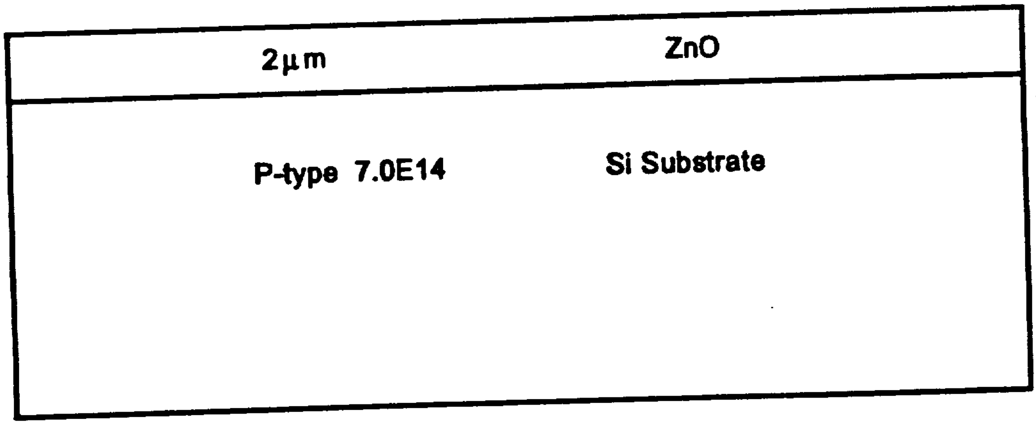
2µm
ZnO
P-type 7.0E14
Si Substrate

Figure 2.4.4 Material layout for ZnO overlay on a silicon substrate.

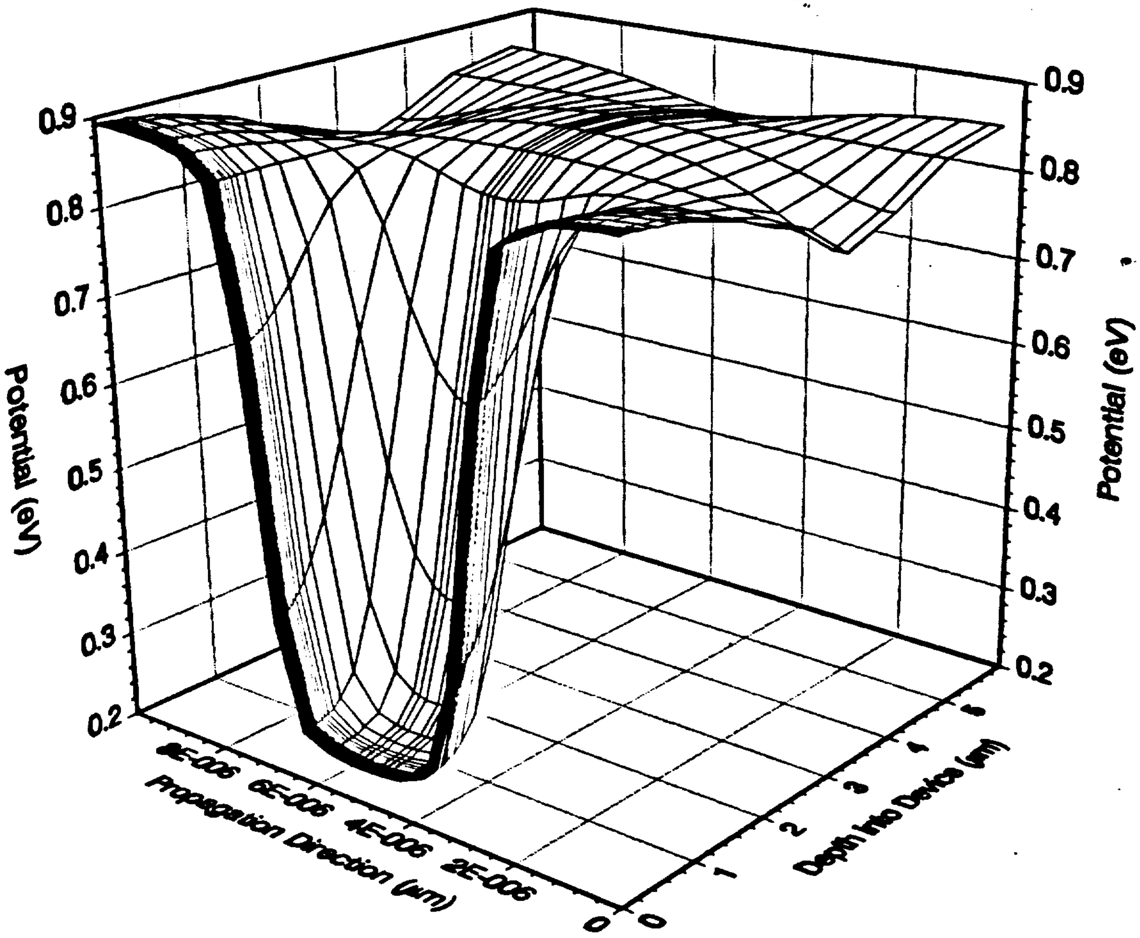

Figure 2.4.5 Conduction band equi-energy contours in the silicon indicating a potential well which is capable of transporting charge.

2.5 Materials Growth

In previous reporting periods materials growth development has focused on the development of an indium-free mounting technique for 2 inch substrates along with investigation of the spatial uniformity of dopant and host fluxes. In the present reporting period efforts have concentrated in three areas; substrate heater modification, system cleaning and temperature controller upgrading.

2.5.1 Substrate Heater Modification

Investigations into the substrate temperature uniformity showed that large temperature variations across the substrate surface were present. These temperature variations were due to the four filament design of the substrate heater in the Varian Gen II molecular beam epitaxy system. Even upon rotation and the incorporation of a pyrolitic boron nitride diffuser between the heater elements and the substrate rear surface, the effects of temperature variations could not be eliminated. Therefore, a completely new substrate heater was designed, tested and incorporated into the existing substrate heater mount. This heater is based upon a broad area pyrolitic boron nitride encapsulated graphite heater element. Not only has this heater lead to a much more uniform substrate temperature, but is also mechanically more robust than the original Varian design. This design should eliminate substrate heater failure, which in the past has been the major reason for system down time.

2.5.2 System Super Cleaning

The second major effort in the past six months has been the complete cleaning of the system. It was determined that the surface particulate density was higher than acceptable for the large area devices required for this program. The majority of these particulates were found to be generated in the growth chamber. To reduce the particulate density the entire system was opened and scraped out. This has resulted in an order of magnitude improvement in the system base pressure, along with a shortened bakeout time. A reduction in surface particulate density by a factor of four has be observed after the system cleaning.

2.5.3 Gallium and Aluminum Controller Replacement

Recent publications along with our previous investigations lead us to closely monitor our Al and Ga effusion cell temperatures and fluxes. Shown in Figure 2.5.1 is a typical plot of the gallium effusion cell temperature as a function of time with the original Eurotherm model 919 temperature controller and model 932 SCR. One startling feature is that the temperature undergoes periodic variations with a period of about 400 seconds. This is similar to that which has been observed in other research laboratories. Not only are these periodic variations in temperature observed, but long term variations in temperature of up to +/- 0.5 °C are also observed.

To improve the temperature and therefore host element flux stability the older Eurotherm temperature controllers and SCRs were replaced on the aluminum and gallium cells. A plot of the typical time dependence of the of the gallium effusion cell using the new Eurotherm model 818s temperature controller and model 832 SCR is shown in Figure 2.5.2. As can be seen the

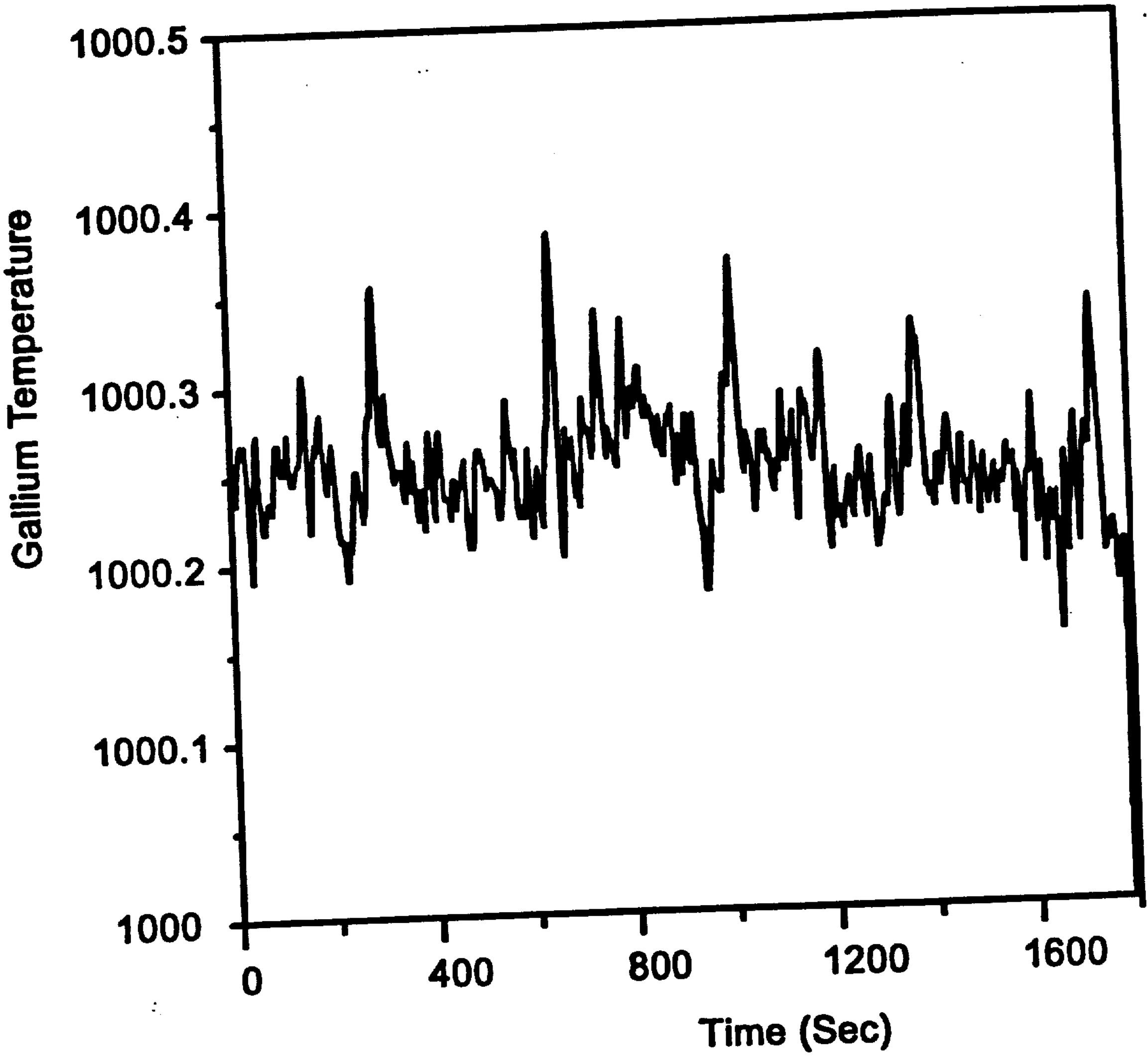

Figure 2.5.1 Gallium effusion cell temperature as a function of time using old controllers.

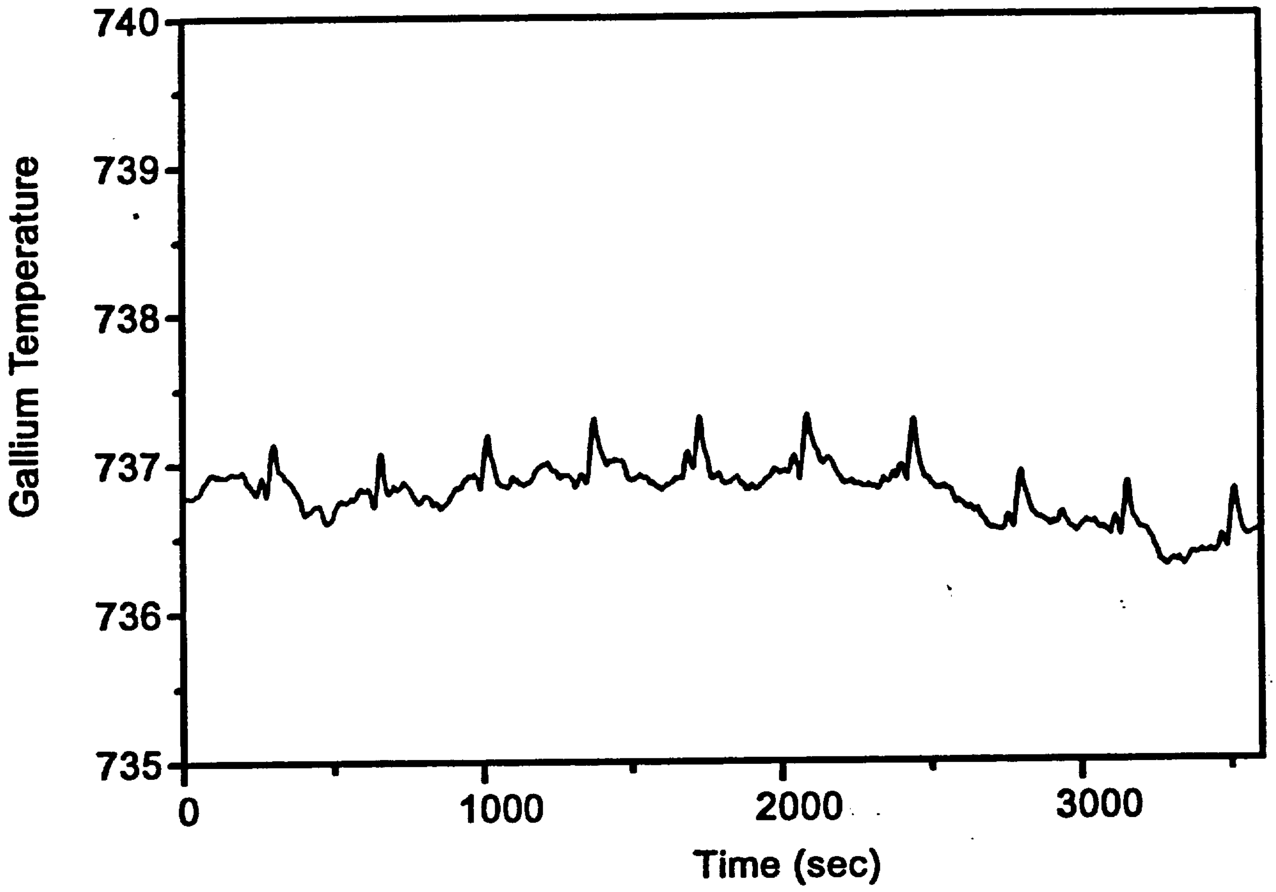

Figure 2.5.2 Gallium effusion cell temperature as a function of time using new controllers.

periodic temperature disturbances have been essentially eliminated and the total variation in temperature has been reduced to +/- 0.1 °C. This has resulted in a very stable effusion cell flux as shown in Figure 2.5.3 where the time dependence of the aluminum cell flux, as monitored by an ionization gauge which has been rotated into the growth position, is shown over a period of 1800 seconds. The total variation of the flux has be reduced to less that 2 %.

Another effect leading to nonuniform doping and/or growth rate that has been investigated is the flux transients that are caused by changes in effusion thermal loading when the effusion cell shutters are opened to initiate growth. The aluminum flux as a function of time is shown in Figure 2.5.4, as the shutter is opened. As can be seen during the first 200 seconds of growth (~ 450 A) the aluminum flux decreased by 14%. Not only are the flux transients observable upon the initiation of single layer growth but are also seen during superlattice growth. The gallium flux as a function of time during the growth of a GaAs/AlAs superlattice structure is show in Figure 2.5.5. In this growth sequence the Ga shutter was open for 100 sec. and then closed for 100 sec. Upon each shutter opening in the structure a flux transient is observed. These flux transients lead not only to slight inaccuracies in the layers thickness, but also to small variation in the doping densities near the interfaces. At present feed-forward temperature control algorithms are being developed to eliminate these flux transients.

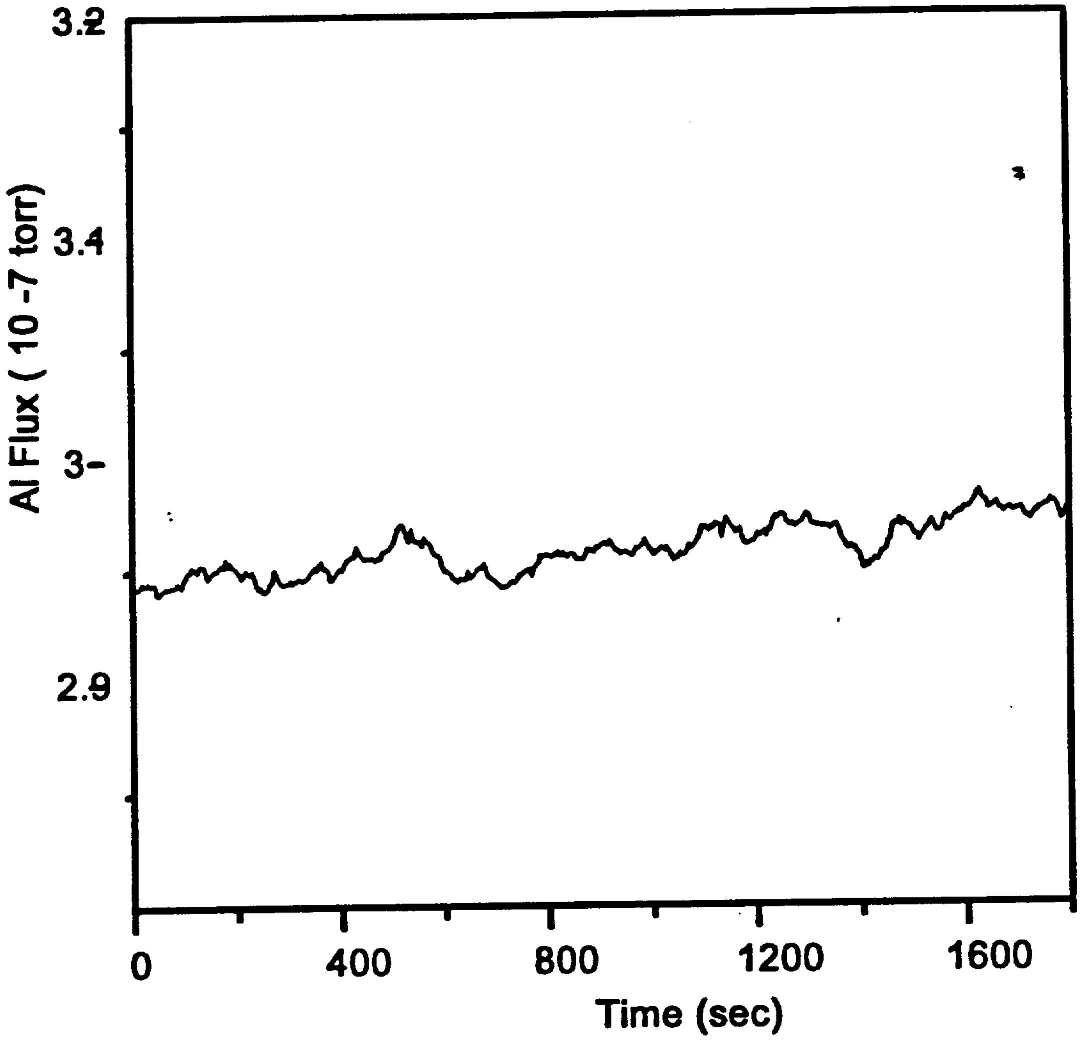

Figure 2.5.3 Stable effusion cell flux obtained using new controllers.

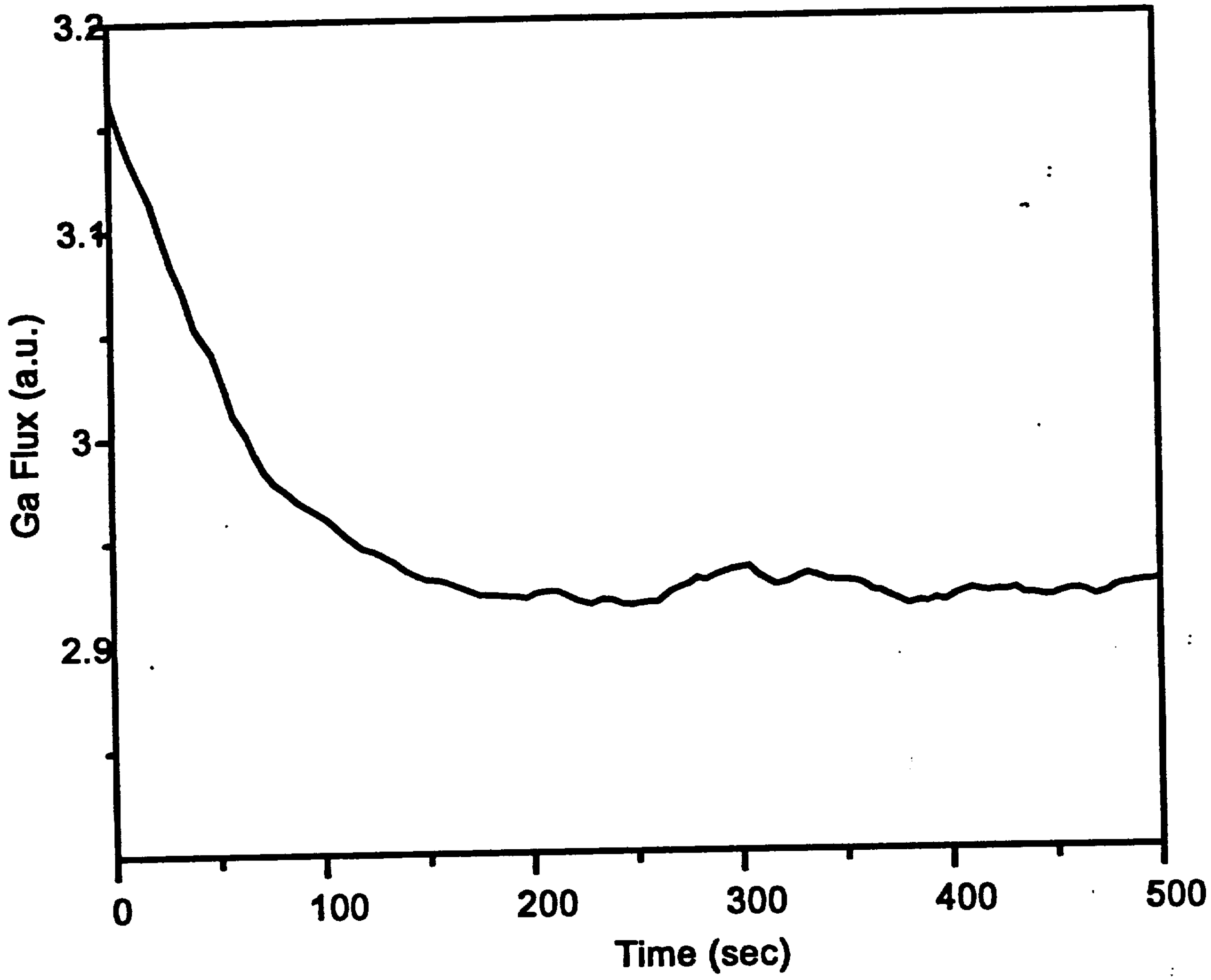

Figure 2.5.4 Aluminum flux as a function of time after initially opening shutters, flux decreases by 14%.

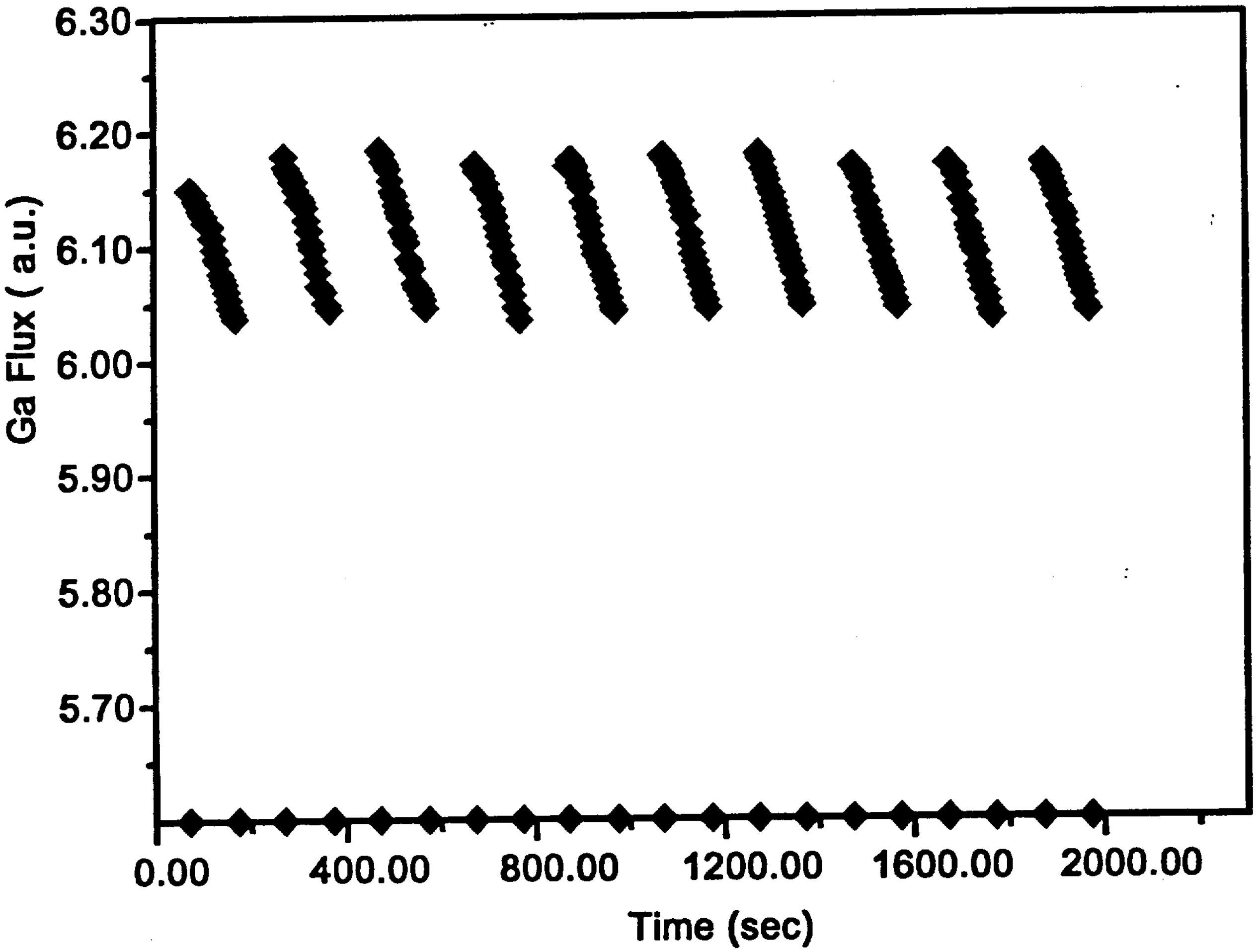

Figure 2.5.5 Gallium flux transients observed when the shutters are periodically opened and shut for 100 sec intervals.

3. Charge Transfer and Overflow Device

3.1 Introduction

Of the three devices that comprise the acoustic charge transport (ACT) imager, the charge transfer device (CTD) plays a major role in its operation, where the CTD collects and stores the electrons generated in the avalanche photodetector (APD), then transfers this accumulated charge to the ACT device for readout. The charge storage and overflow properties of the CTD directly affect image quality, while the transfer capabilities in part determine the speed and fidelity of image readout.

In this section the progress made on the CTD will be presented. An extensive review of charge storage device technology is presented, along with its application to the CTD. The electrical characteristics of the earlier vertical CTD (V-CTD) structures have been explained in light of other GaAs charge storage devices, as will be described to follow. This work on the V-CTD is also relevant to the development of the lateral CTD (L-CTD), since vertical charge confinement is necessary in either architecture. This section will conclude with future goals for CTD research and imager development.

3.2 Review of Charge Storage Devices

3.2.1 General Concepts for Charge Storage Device

From a review of Si CCDs [3.2.1], two modes of charge storage are used in solid state imaging devices, the surface channel or enhancement approach and the buried channel or depletion approach. The charge storage operation in either mode is the same; they only differ in how the charge storage region is created. The features common to both modes are introduced first, followed by a description of the different charge storage regions in the surface channel and buried channel devices. In the following discussions, the stored charge could be either electrons or holes, however, only electron storage will be described since the APD will operate with electron multiplication.

In common to both storage modes, there exists a steady state concentration of electrons in the charge storage region or well of the device, where this electron concentration represents the full well capacity of the storage device. In this steady state, full well condition the device is not necessarily in equilibrium, i.e. it may be biased. Through the application of appropriate biases, a non-steady state, transient condition is created where the electrons in the storage region are removed to yield an empty well ready for charge storage. Electrons can be introduced into the storage well either by lateral injection or through optical generation within the well, until the electron concentration reaches its steady state value. Since the storage device is operating under transient conditions, the storage well will eventually return to the steady state, full well condition even if no electrons are purposely introduced. This degradation of the charge storage well occurs either by leakage of charge into the charge storage region, such as the reverse bias leakage across a p-n junction, or via spurious generation of electron - hole pairs, possibly aided by defects, within the well. These so-called dark current mechanisms determine the maximum storage time for charge in the device.

Charge storage devices differ in how the potential well is created. In the surface channel or enhancement mode device, for example a Si MOS capacitor, charge is stored in an inversion layer at the interface between a semiconductor and insulator. This inversion layer, in turn, is surrounded by a depletion layer which acts to confine the electrons both vertically and laterally. The charge capacity of the surface channel device is determined by the bias applied to the gate and by the capacitance of the dielectric. In a buried channel or depletion mode device, charge is stored in a potential well vertically defined by two space charge regions. These space charge regions can be produced by p-n junctions, an MOS structure, or a schottky barrier. An example of an depletion mode device is a GaAs MESFET structure, where the schottky gate contact has been biased to pinch-off the channel. Table 3.1 shows some general advantages and disadvantages for these two charge storage modes.

Table 3.1: Comparison of Charge Storage Modes for Solid State Imaging Devices.

	ADVANTAGES	**DISADVANTAGES**
Enhancement mode	• largest possible charge storage capacity • simpler variation between Q and C compared to depletion mode • devices can be tested with no lateral isolation mechanism, such as a junction or mesa, because of depletion layer surrounding the inversion layer	• need the proper insulator - semiconductor interface, none easily implemented currently for GaAs • slower charge transfer due to interface states • other standard problems associated with MOS devices in Si due to presence of oxide
Depletion mode	• doesn't require a good insulator - semiconductor interface. can use Schottky barrier or p - n junctions • fast charge transfer since low surface states • can use positive, negative, or both bias schemes • more radiation tolerant	• leakage current often limits storage time rather than generation; especially true for GaAs schottky barriers • smaller charge capacity compared to enhancement mode • may require an extra contact to pinch-off or totally deplete structure • need device isolation for test structures, probably mesa or schottky barriers

3.2.2 Charge Storage Devices in GaAs

As noted in the previous section, there is no suitable insulator for GaAs that would allow enhancement mode charge storage devices based on a GaAs - dielectric interface, although progress is being made using novel structures grown by MBE [3.2.2]. Therefore, GaAs charge storage devices, such as CCDs, are based on the buried channel approach [3.2.3]. The standard GaAs CCD is based on a MESFET structure on a semi-insulating substrate, and is typically used for high speed signal processing and benefits from the easy integration of MESFET based circuitry with the CCD.

An important consideration for the CTD is the storage time of the device. As discussed in the previous section, the mechanisms responsible for dark current degradation of a charge storage time are generation in the depleted well, which for GaAs should be a slow process because the generation rate is proportional to n_i, and leakage currents. In such devices, leakage is the main well degradation mechanism, because of the small barrier height for schottky barriers on GaAs coupled with the small generation rate for GaAs. To assess the maximum leakage current tolerable by the CTD for HDTV applications, a simple calculation has been made following Fossum et al. [3.2.4], assuming that the storage time is leakage-limited, the maximum storage time necessary for the CTD is 5 ms (170 frames per second), and the well capacity is 10^6 $e^- cm^{-2}$, a value consistent with current Si CCD-based HDTV imagers. This calculation yielded a maximum leakage current of 1.6 $nAcm^{-2}$. This value should be compared with the lowest reported leakage current reported by Fossum et al. for 2DEG-CCDs of 30 μAcm^{-2} and a Rockwell result for a MESFET structure using an undoped AlGaAs layer of 0.01 - 0.1 $nAcm^{-2}$. Although the Rockwell value appears encouraging, their results were for large area devices, where surface effects are less significant. For the necessarily small area devices needed in the HDTV imager, such low leakage currents may not be possible. Therefore, an important issue that must be addressed for the CTD is development of low leakage current structures.

GaAs CCDs are based on both standard MESFET and high speed HEMT (2DEG-CCD) structures [3.2.4]. Design constraints represent a trade-off between storage capacity and transfer efficiency, where increasing the storage capacity yields a large potential trough between the wells which captures signal charge during transfers. This problem is not as severe for Si CCDs because of the use of oxide passivation between the wells and overlapping gate structures. The minimum energy of this potential trough between wells increases with increasing channel doping and decreasing depth of the well from the surface, and decreases with decreasing well separation. Typical storage capacities for MESFET CCDs are around $1 - 5 \times 10^{11}$ cm^{-2}, while 2DEG-CCDs have larger capacities around 1×10^{12} cm^{-2}. However, because the channel in a 2DEG-CCD is located very near the surface, the charge transfer efficiency suffers unless the wells are quite close (~ 1 μm). This factor may not be important for application of 2DEGs to CTDs, since only one transfer is necessary between the CTD well and the HACT channel, and resistive gate technology could be used to remove the inter-well potential troughs [3.2.4]. A greater limitation to the use of 2DEGs for CTDs is the large leakage current of the schottky barrier gate due to the high electric field caused by the n^+ AlGaAs layer [3.2.4], an issue that can be addressed through the use of p-i-n doped 2DEGs or improved schottky barrier technology.

Charge storage devices have also been developed which are based on depletion mode storage using p-n junctions, where very long storage times have been observed [3.2.5, 3.2.6]. However, in these devices no free charge is transferred between storage wells, but rather only a capacitance change due to the depleted charge is sensed, for use as digital memory devices. Therefore, these structures are not directly applicable to the CTD.

Although a suitable insulator is not readily available for GaAs, enhancement mode devices have been attempted utilizing the heterojunction discontinuities between GaAs and AlGaAs, where the AlGaAs barrier is undoped. This structure is used in the heterostructure insulator gate FET (HIGFET) device. Unfortunately, because of the small energy barriers at the GaAs - AlGaAs interface ($\Delta E_C = 0.263$ eV, $\Delta E_C = 0.132$ eV for x = 0.3 AlGaAs) [3.2.7], the gate leakage currents are very large, in low milliampere range at room temperature [3.2.8], which would result in very small storage times. Kleine et al. [3.2.9] at Purdue have examined enhancement mode operation of an GaAs - AlGaAs structure, and only observed measurable electron inversion charge below 70 K, where charge loss was limited by thermionic emission across the heterojunction conduction band discontinuity.

One other GaAs storage approach that has been investigated is a floating gate device, where charge is stored in an GaAs - AlGaAs quantum well [3.2.10]. However, some question exist about the mechanism for charge storage in these devices, and that trapping effects may be responsible [3.2.11].

3.3 Results

3.3.1 Comparison Between Simulations and Experimental Results for CTD-5

The capacitance-voltage (CV) and transient capacitance (Ct) data from the V-CTD structure CTD-5 has been described based on simulations using Atlas II and HETMOD, a device simulator made available to Georgia Tech by IBM. Specifically, the storage capacity for CTD-5 has been calculated and aspects of the CV and Ct data explained. Simulation are currently being performed on CTD-8, with results to be presented in the next report. This work has been hampered by the high failure rate of CTD-8 devices, as will be described in section 3.3.2.

The equilibrium band diagrams for the two different V-CTD device structures are shown in Figures 3.1 (a) and (b), where we assumed originally that the charge storage well was formed by the AlGaAs layer and either a volume-doped (CTD-5) or delta-doped (CTD-8) p-type layer.

To determine if charge storage could occur in the well, the effect of bias on the electron concentration in the well was simulated for CTD-5. For charge storage to occur, the free electrons must be removed from the well. The integrated electron density in the well was 2.7 x 10^{12} cm^{-2} in equilibrium. However, no change in electron density, either an increase or decrease, was observable for readout (surface) biases from -1.0 to 1.50 V. Therefore, the well formed by the AlGaAs layer and the p barrier layer does not function as a charge storage well. The lack of charge modulation in this region results from all the applied bias being dropped across the two depletion layers, so that there is no potential variation and thus no charge modulation within this well.

However, charge storage was observed in CTD-5 from the Ct measurements described in previous reports. The question is where in the structure is charge being stored? We can understand the charge storage mechanism operating in CTD-5 from an analogy to the GaAs MESFET depletion mode CCD, where an n-type channel is depleted by a suitable bias to pinch-off. In CTD-5, this channel is the p-type barrier region, with its barriers formed by the depletion layers of the two p-n junctions. During a Ct measurement, a bias is applied to the readout which depletes holes from the p-type region; when this bias is removed the p-type region fills with holes

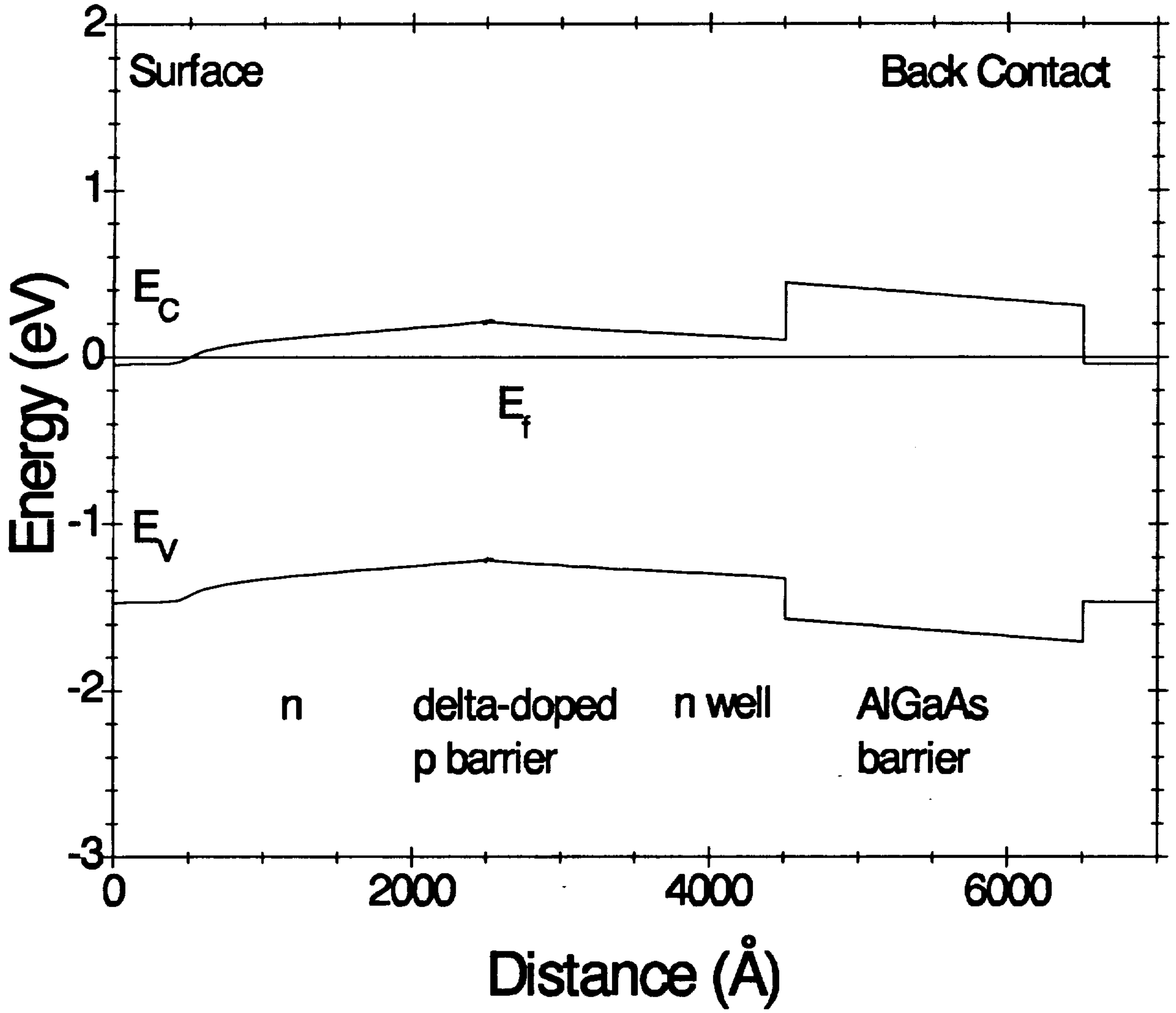

Figure 3.3.1 (a) Equilibrium band diagram of a delta doped CTD structure (CTD-8).

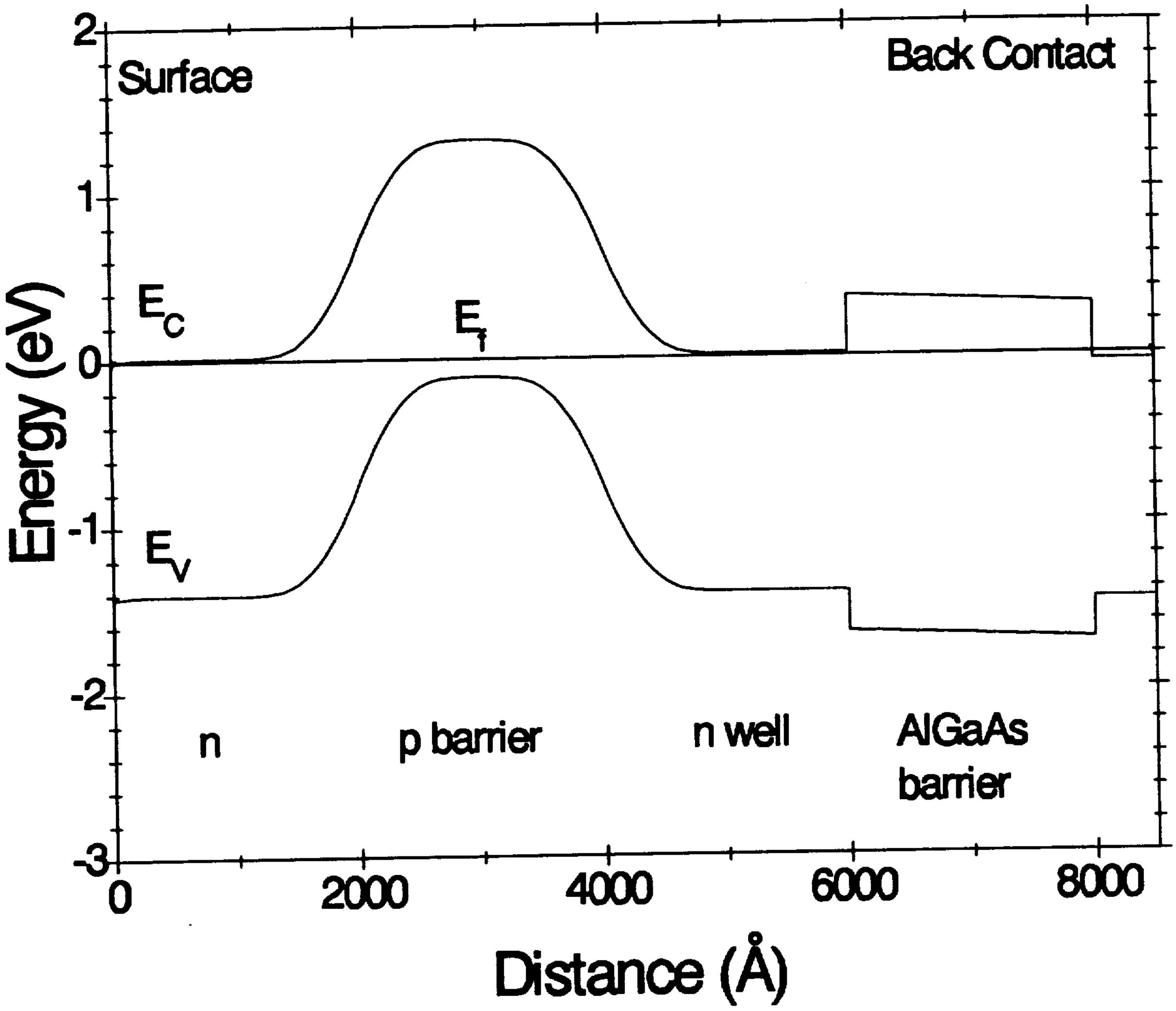

Figure 3.3.1 (b) Equilibrium band diagram of a volume doped CTD structure (CTD-5).

via generation and leakage across the p-n junctions. Therefore, this structure acts as a hole storage device. Using HETMOD a pinch-off voltage and hole storage capacity for CTD-5 were calculated of 1.49 V and 1.15 x 10^{12} cm^{-2}, respectively.

The storage lifetime (1/e value) observed in CTD-5 from Ct measurements at bias pulses slightly less than required for breakdown (± 4 V) were $\tau_s = 3$ ms. The lifetime is too short for the CTD, and is also small compared to literature values on the order of minutes, which indicates that either the material has a large number of defects which increases the generation rate in the p-type region or that leakage currents across the p-n junction barriers are large. From previous work we had identified that the leakage currents for CTD-5 were sizable and resulted from shunting of the front n-p junction due to excessive alloying from the AuGe - based ohmic contacts. Therefore, with the improved PdGe contact technology the charge storage time should increase.

3.3.2 Device Characterization of CTD

Investigation of the failure modes for V-CTDs using CV and conductance-voltage (GV) measurements have been performed as a function of device area, which isolates the effect of bulk phenomenon from mesa sidewall surface effects on device breakdown. For the volume-doped CTD-5, device failure resulted from ohmic contact diffusion into the junction region, as described previously. Therefore, an alternative ohmic contact technology for the V-CTD was developed based on PdGe metallization, which was used in all subsequent devices. For the delta-doped CTD-8, the major contributor to breakdown was a bulk-related effect, which we believe resulted from electric field induced failure of the Be delta-doped barrier layer due to movement of the dopant ion [3.3.1]. Future delta-doped structures will utilize low temperature, As rich growth conditions to minimize Be diffusion, thereby improving the lifetime and operating characteristics of these devices.

Initial investigation have begun on the effect of traps on CTD operation. The variation in capacitance with test frequency is being studied, where traps are not expected to contribute to the measured capacitance with increasing frequency, thus causing a decrease in capacitance.

Storage lifetime measurements have recently been performed in liquid nitrogen (LN_2) at 77 K, to study the effect of temperature on leakage currents and generation rates. Significant enhancements in the storage lifetime for CTD-5 have been observed, with a value of 68 s for a - 1 V bias pulse at 77 K, compared to 2.9 ms for a - 4 V bias pulse at 300K. Future work will examine the dependence of storage lifetime on the bias pulse, pulse duration, and temperature, as a precursor to deep level transient spectroscopy setup and testing.

3.3.3 Fast Storage Lifetime Testing System for CTD

An approach for determining the very short charge storage lifetimes, of the order of microseconds, is to implement a fast transient testing system based on voltage sensing, rather than capacitance sensing, since the minimum capacitance measurement time is at best 10 μs, compared to voltage measurements on a digital oscilloscope with times less than 1 μs. In such an approach a voltage proportional to the RC time constant for the CTD is measured after a pulse bias, and the dependence of device capacitance on time is deconvoluted from the time-dependent RC product. This approach actually determines a dc capacitance, as opposed to a standard capacitance meter which measures an ac capacitance.

Such a fast voltage transient system has been modeled at GTRI using the circuit simulation program PSPICE, to define a test setup where the CTD capacitance and a known resistance determine the RC time constant of the measurement setup, rather than any parasitic capacitances and resistances associated with the setup. This system has recently been setup, and is now being tested at GTRI.

3.3.4 Summary and Future Work

Significant progress has been made in our understanding of the operation of previous CTDs. This knowledge will allow us to more effectively design device structures using the HETMOD and Atlas II simulators for subsequent growth , fabrication, and testing. From the review of GaAs charge storage devices and the experimental results on CTD-5, an important need for long storage time CTDs is a low leakage current device structure, which can be realized by a high quality GaAs - insulator interface. Such a structure would benefit both depletion mode devices by reducing leakage currents, and also allow the fabrication of enhancement mode devices. Such a technology would also find application in many other areas of GaAs device technology. Work will also begin on construction of a deep level transient spectroscopy system for enhanced characterization of CTDs, APDs, and general MBE grown material.

4.0 Avalanche Photodiode Development

4.1 Theory of Impact Ionization and Modeling of Avalanche Photodiodes

One of the keys to developing an accurate model of an avalanche photodiode for device simulation is to formulate the impact ionization transition rate correctly. Most early theoretical studies of the impact ionization rate have limited validity since they neglect the complexity of the phonon scattering processes and the energy band structure of the host material. Numerical transport models based on the ensemble Monte Carlo method have been more recently employed [4.1.1-4.1.3] which include the full details of the band structure based on either a pseudopotential or $\mathbf{k} \circ \mathbf{p}$ calculation. The impact ionization transition rate is typically formulated within these numerical transport models using a parameterized expression called the Keldysh formula [4.1.4]. The Keldysh formula contains two parameters, p and E_{th}, which are very difficult to determine from first principles. Both of these parameters are selected such that the calculated ionization rate agrees well with experimental measurements. Often, more than one set of parameters can reproduce the same ionization rate in a given material. Subsequently, the nature of the ionization process cannot be uniquely determined from use of the Keldysh formula alone. In addition, the Keldysh formula is derived assuming several simplifying assumptions. These are that the transition is calculated assuming a direct semiconductor with parabolic bands for all of the carriers, no k-dependence is considered, and the overlap integrals are constant for all initial energies.

There have been some recent attempts by several investigators to address the limitations imposed by the use of the Keldysh formula [4.1.5-4.1.9]. During the course of this project, we have developed a completely numerical study of the interband impact ionization transition rate for both bulk silicon and GaAs [4.1.8, 4.1.9]. The ionization transition rate is evaluated using a wavevector dependent, semiclassical formulation of the transition rate. The transition rate is determined using Fermi's golden rule from a two-body screened Coulomb interaction assuming energy and momentum conservation. The overlap integrals in the transition rate expression are determined through use of numerically generated wavefunctions determined via a **k**∘**p** calculation. The full details of the calculation and the model are reported in references [4.1.8] and [4.1.9] which are enclosed within the appendix. It is found from these calculations that the transition rate is greater for electrons originating within the second conduction band than the first conduction band and that the threshold can be classified as relatively soft in both GaAs and silicon. It is further found that the threshold is somewhat harder in silicon than in GaAs.

Though the numerical treatment of the interband impact ionization transition rate is useful it is highly computationally intensive. As such calculations of the transition rate typically require many cpu hours. In an attempt to reduce the computational demands necessary for a completely numerical formulation of the transition rate, we have investigated the usefulness of an alternative analytical formulation of the transition rate [4.1.10]. The new wavevector dependent analytical formulation was first developed by Quade et al. [4.1.11]. The details of their model have been sketched out in reference [4.1.10] which is enclosed within the appendix. It is found that the wavevector dependent model of Quade et al. produces very similar results to those obtained with the completely numerical model for some quantities. Specifically, both models predict that the effective threshold for impact ionization in both GaAs and silicon is quite soft, that the majority of ionization events originate from the second conduction band in both materials, and that the transition rate is wavevector dependent. Nevertheless, there are significant limitations to the usage of the Quade formula for modeling interband impact ionization. These are that the Quade formula does not accurately reproduce the quantum yield data for bulk silicon, it requires determination of a new parameter, related physically to the overlap integrals of the Bloch states, and fails to account for any wavevector dependence of the overlap integrals themselves. As such the transition rate can be seriously overestimated at those points within the Brillouin zone for which nearly vertical, small change in wavevector, transistions occur. Based on these results, we have concluded that the Quade formula may not be truly useful in avalanche photodiode simulation and have retained the completely numerical technique for use in the next generation of APD device simulators.

In addition to our work on developing an improved theory of interband impact ionization, we have further worked on refining our APD device simulator as well as work on studying the performance of APD devices. Towards this end, we have begun work on tailoring our existing hydrodynamic simulator to study APD structures. This work has progressed along two related directions. First, we have fashioned the simulator for studying cylindrical geometries. This has been used to study a separate absorption, multiplication APD used by Bell Northern Research for lightwave communications systems. In addition, we are presently investigating the inclusion of nonparabolicity effects into the hydrodynamic model to greatly enhance its accuracy for high energy devices. Two different formulations are being investigated, a simple nonparabolicity model and a power law model. Though the power law model offers somewhat better accuracy at high energy, its use is somewhat frustrated by the need to evaluate some of the moments numerically. We expect that we will soon incorporate the nonparabolicy technique into our existing hydrodynamic simulator enabling its usage for APD device simulation and development.

4.2 Characterization of Avalanche Photodiodes

During this report period, a paper entitled "Gain Properties of Doped GaAs/AlGaAs Multiple Quantum Well Avalanche Photodiode Structures" was submitted and accepted for publication in the Journal of Applied Physics Letters. The paper presented experimental confirmation of a structure-induced carrier multiplication in the low-voltage regime due to interband impact ionization in the MQW APDs. This was contrasted with data from a conventional p-i-n APD where no such gain existed. The details of the experiments and the results were described in the previous report.

Recent work has concentrated on performing several device simulations to help gain some insights into the physical phenomena that take place in p-i-n's, undoped, and doped MQW structures. Various simulations using the Gummel and Newton's methods were used to calculate electric field distribution, valence and conduction band potentials and carrier concentrations, as well as electron and hole mobilities throughout the structure.

Figures 4.2.1(a) and 4.2.1(b) show the electric field profile and band potentials for a p-i-n structure. Both the p- and n- layers are heavily doped (1×10^{18} cm^{-3}), while the i-layer is made up of intrinsic GaAs. Note the two peaks in the field plot which are the result of band bending at the edges due to bandgap narrowing caused by heavy doping. Figures 4.2.2(a) and 4.2.2(b) present the carrier concentrations and mobilities throughout the p-i-n structure. Electron and hole mobilities are lower in the doped GaAs regions due to ionized impurity scattering.

Figures 4.2.3 and 4.2.4 show the results of the simulation when applied to an undoped GaAs/AlGaAs MQW structure with four built-in wells, while Figure 4.2.5 show the effect of

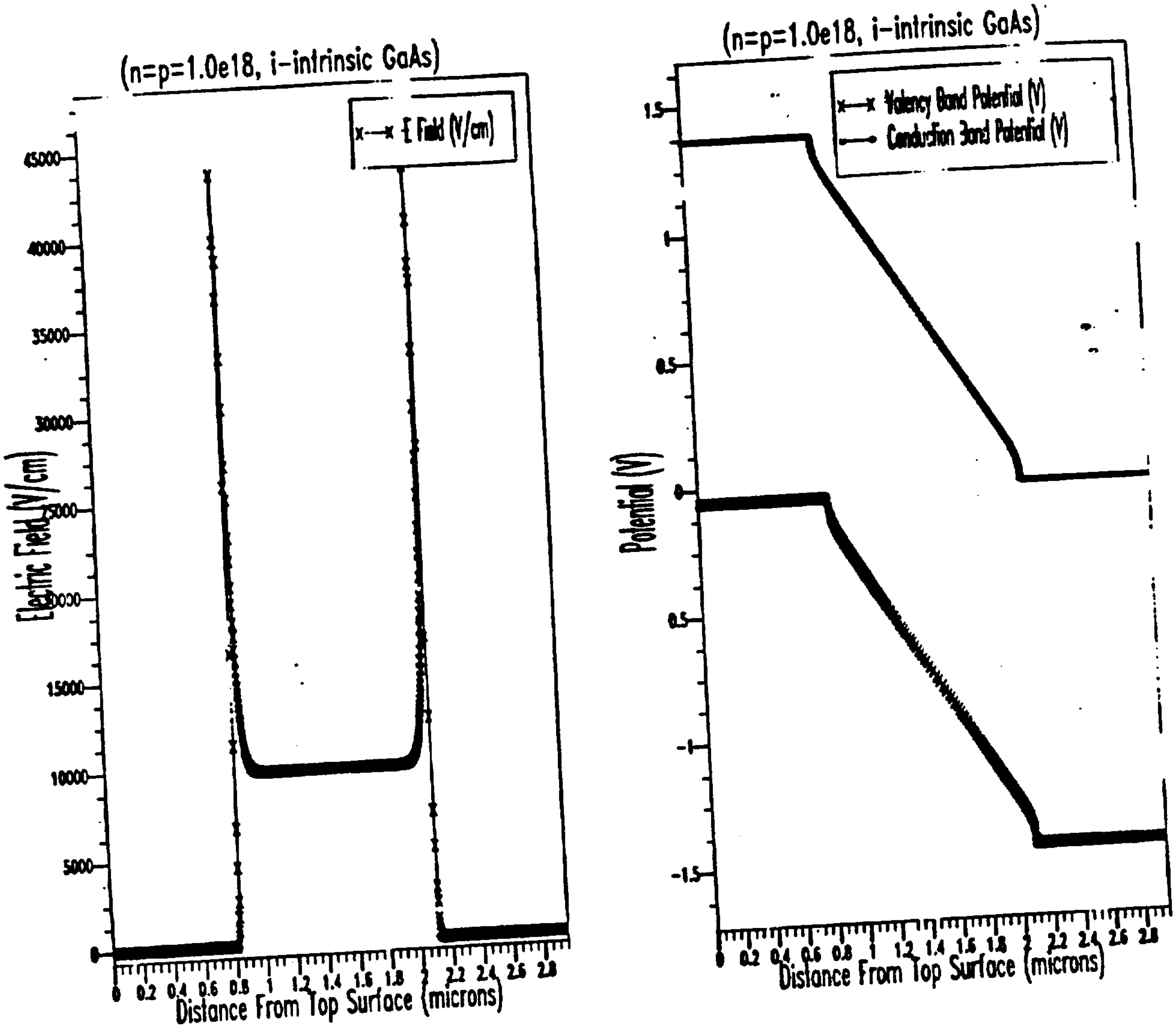

Figure 4.2.1 (a) Electric field profile in a p-i-n photodiode structure. (b) Band diagram for the p-i-n photodiode structure.

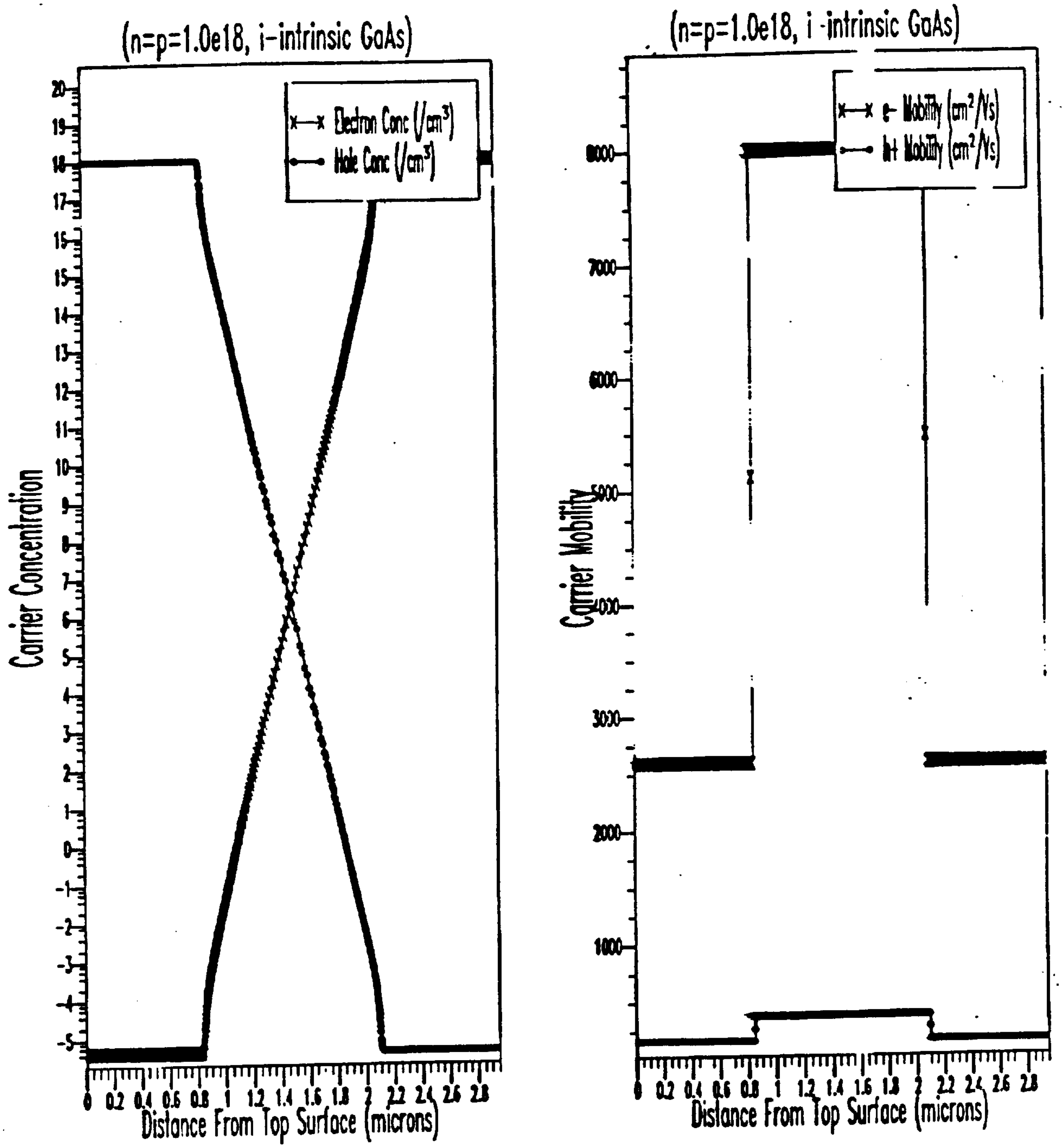

Figure 4.2.2 (a) Carrier concentration as a function of position within the p-i-n structure. (b) Mobility as a function of position within the p-i-n structure.

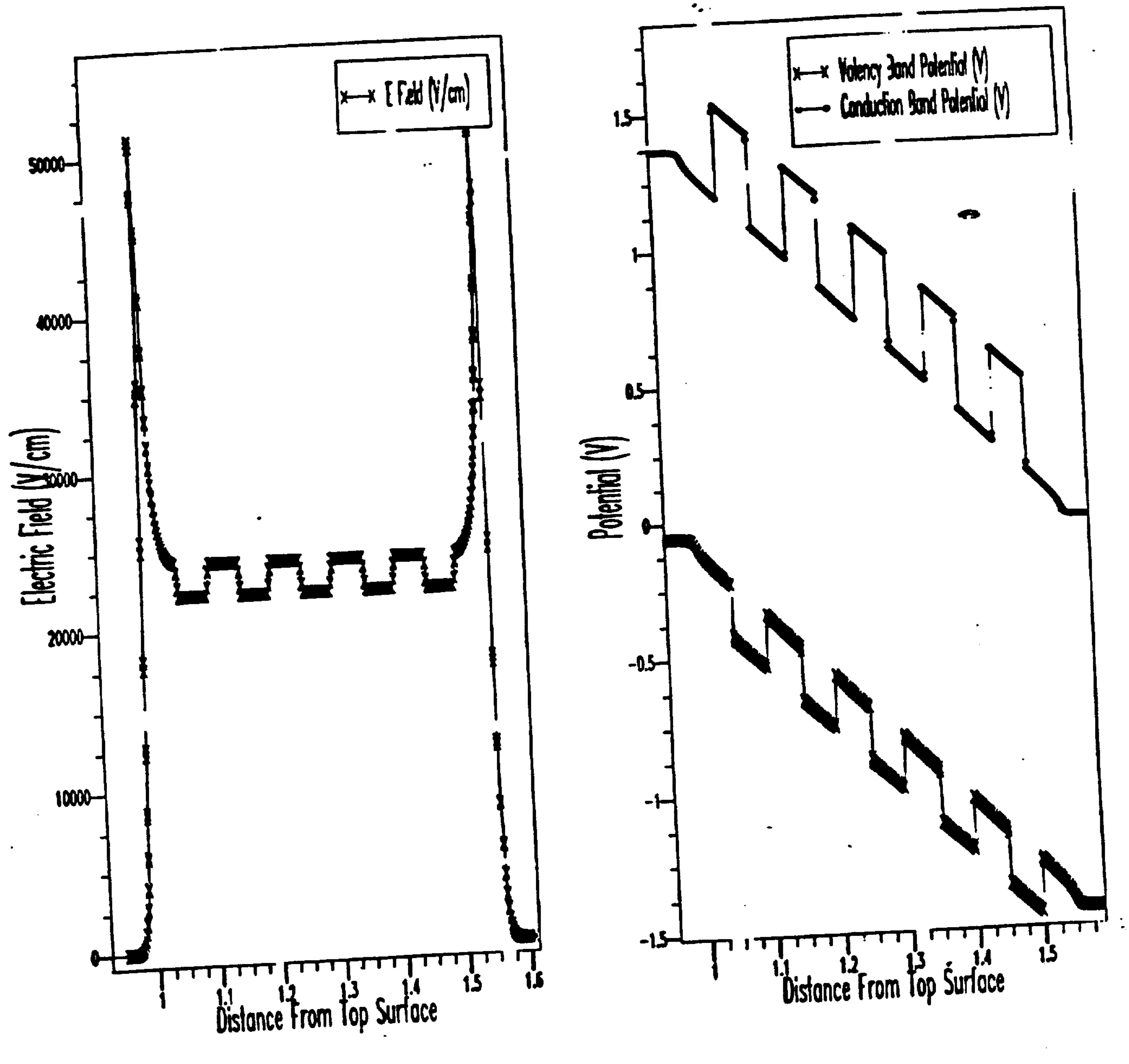

Figure 4.2.3 Simulated electric field and band diagrams obtained for a MQW APD structure.

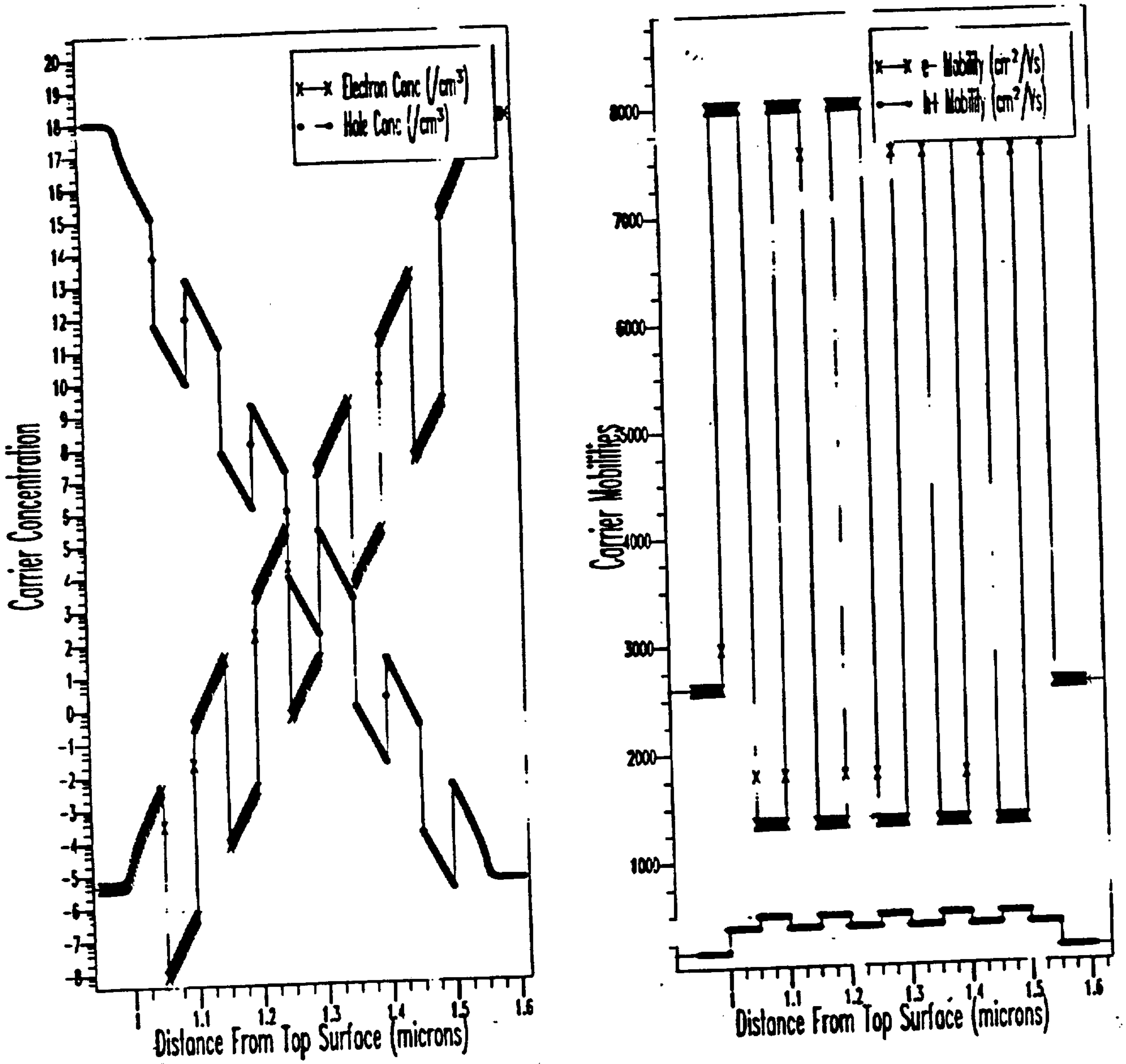

Figure 4.2.4 Simulated carrier concentrations and mobilities as a function of depth obtained for the MQW APD structure.

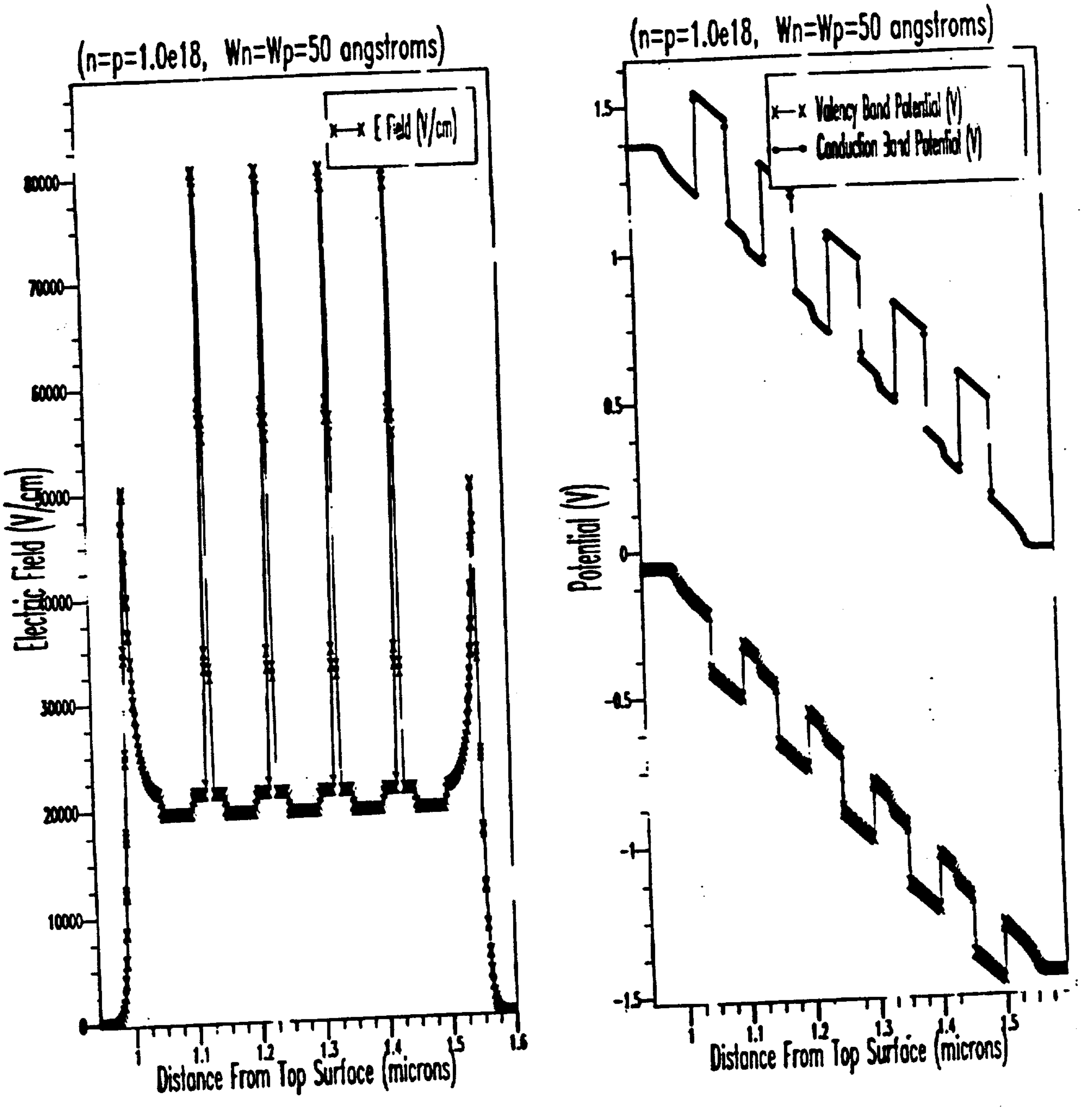

Figure 4.2.5 Simulated electric field and band diagrams for a MQW structure with a 50 Å wide p-n junctions in the center of each GaAs well.

introducing two 50 Å p-n doping layers in the center of every GaAs well. Note, how in Figures 4.2.5(a) and 4.2.5(b), the doping in the wells creates peaks in the electric field and band bending in the potential wells which help increase the kinetic energy of the carriers and thus enhances impact ionization. This clearly illustrates the advantages of using doped MQW structures to greatly enhance the gain characteristics of such devices over that of undoped MQW APDs. The simulation results shown in Figures 4.2.6 and 4.2.7 show a quantitative measure of the relationship between the electron energy gain as a function of the doping density and the doping layer width. In Figure 4.2.6, the p and n doping layer widths were held constant at 50 Å while the doping concentration was doubled (1.5×10^{18} to 3.0×10^{18}). The resulting gain in energy has increased by two fold from about 0.06 eV in Figure 4.2.6(a) to 0.12 eV in Figure 4.2.6(b). In Figure 4.2.7, the same simulation was conducted holding the doping density constant while increasing the p and n layer widths from 100 Å to 150Å. Once again, the energy band offset has doubled from about 0.45 eV in Figure 4.2.7(a) to 0.9 eV in Figure 4.2.7(b) resulting in a substantial increase in the kinetic energy of the electrons.

The simulation shown in Figure 4.2.8 was conducted as an attempt to study the impact of doping imbalances on the field profile and the band structure. Such a problem occurs quite frequently during material growth and is not easily controlled. In this case, we chose the n doping in every GaAs well to be 20% higher than the adjacent doping in the p layer. The resulting electric field profile, as shown in Figure 4.2.9, gradually decreases. This results in partially depleted stages as shown in the band diagram in Figure 4.2.10. Partially depleted and fully undepleted stages in a MQW APD play a major role in limiting the gain characteristics of the device at low bias. Therefore, in order to fully maximize gain at low bias, it is very important

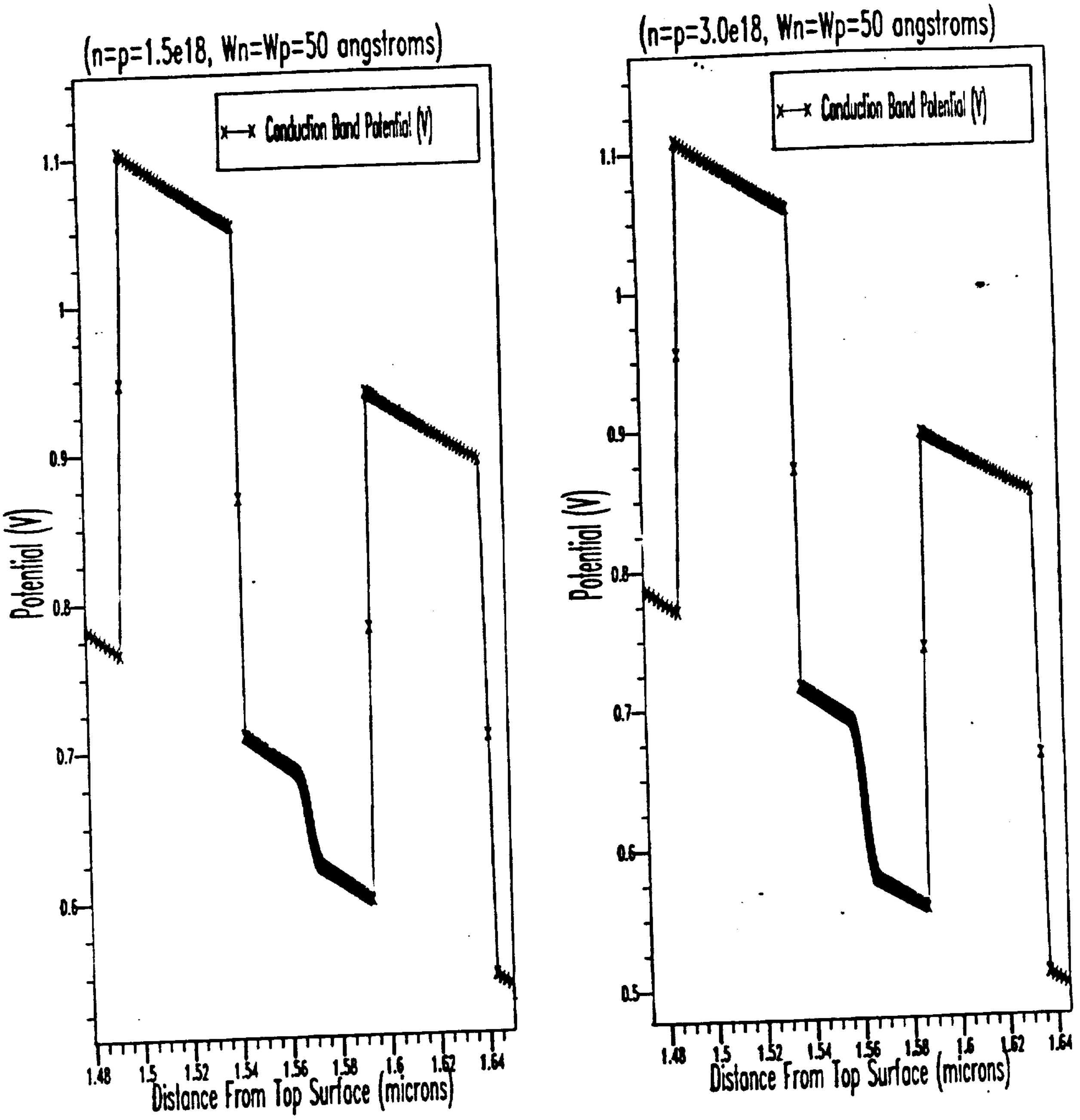

Figure 4.2.6 Energy gain for the p-n junction doped MQW structure when the doping of the p-n junction is altered.

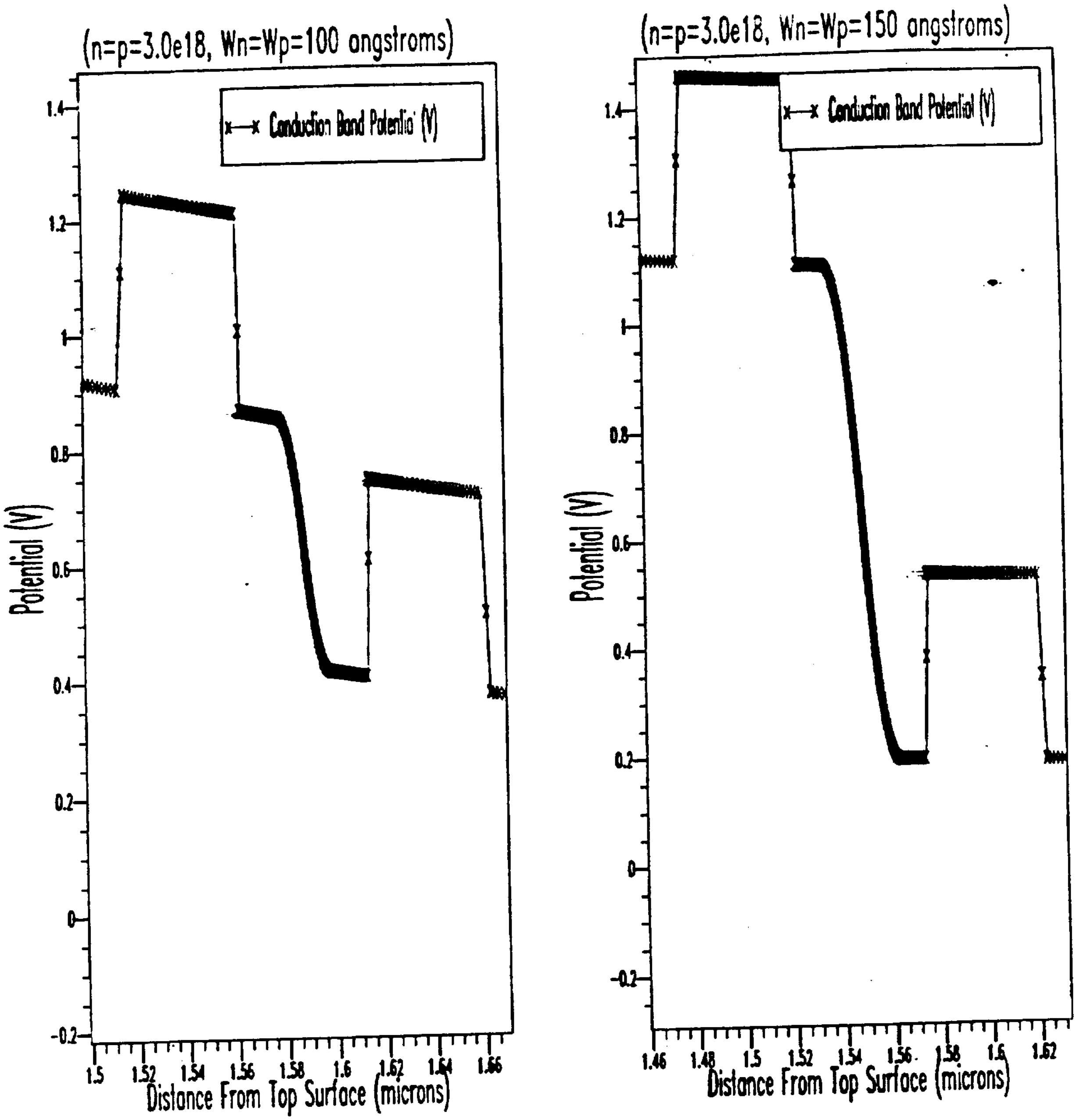

Figure 4.2.7 Energy gain for the p-n junction MQW structure when the thickness of the p-n junctions are increased to 100 and 150 Å respectively.

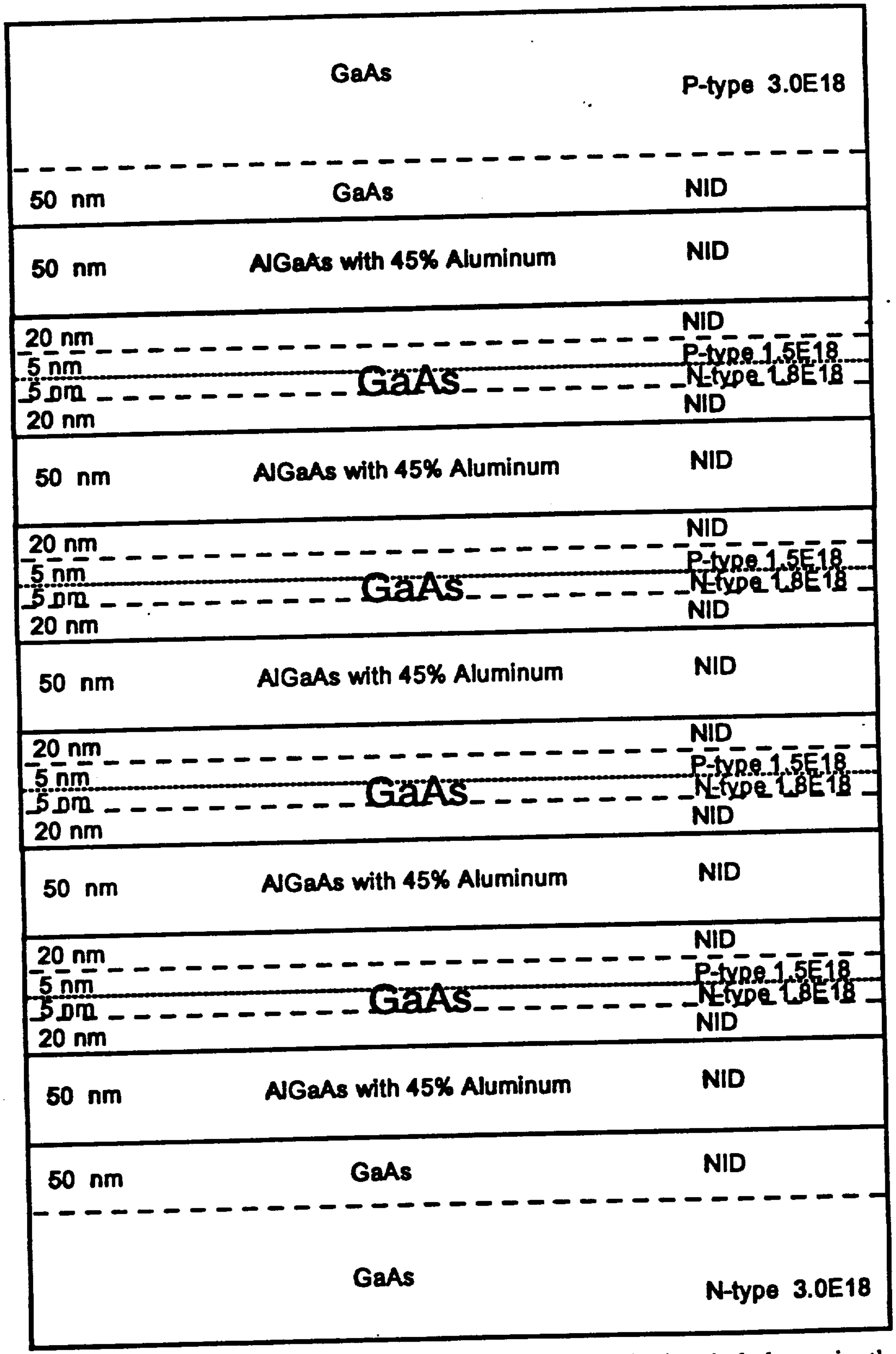
GaAs
P-type 3.0E18
50 nm
GaAs
NID
50 nm
AlGaAs with 45% Aluminum
NID
20 nm
5 nm
5 nm
20 nm
GaAs
NID
P-type 1.5E18
N-type 1.8E18
NID
50 nm
AlGaAs with 45% Aluminum
NID
20 nm
5 nm
5 nm
20 nm
GaAs
NID
P-type 1.5E18
N-type 1.8E18
NID
50 nm
AlGaAs with 45% Aluminum
NID
20 nm
5 nm
5 nm
20 nm
GaAs
NID
P-type 1.5E18
N-type 1.8E18
NID
50 nm
AlGaAs with 45% Aluminum
NID
20 nm
5 nm
5 nm
20 nm
GaAs
NID
P-type 1.5E18
N-type 1.8E18
NID
50 nm
AlGaAs with 45% Aluminum
NID
50 nm
GaAs
NID
GaAs
N-type 3.0E18

Figure 4.2.8 Structure simulated to determine the effect of a doping imbalance in the p-n junction of the MQW structure.

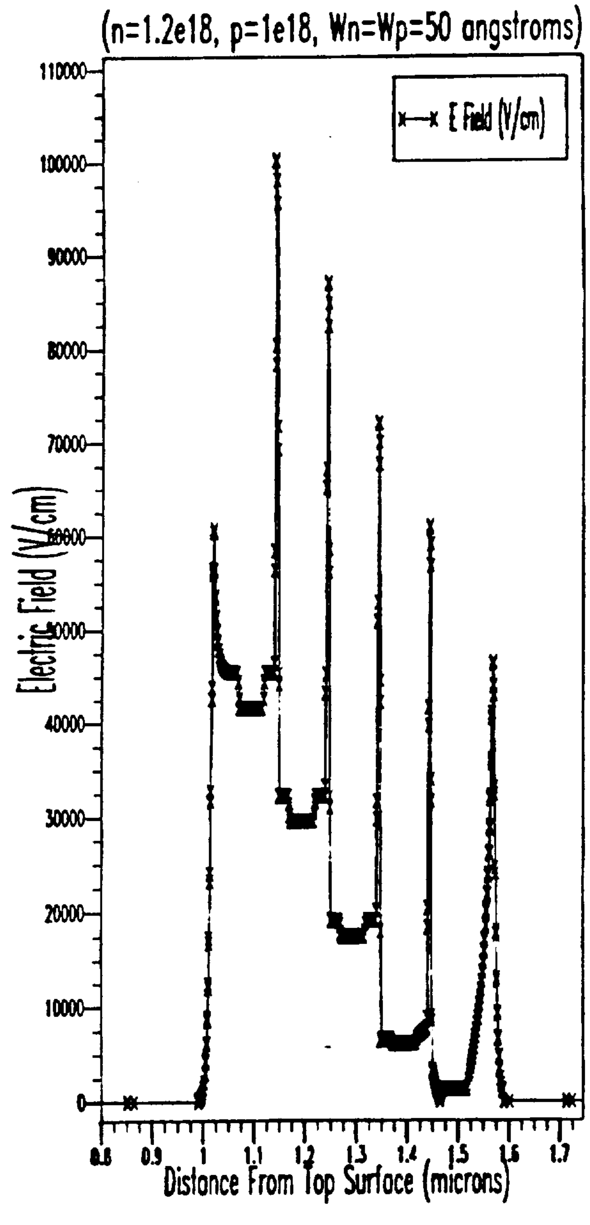

Figure 4.2.9 Electric field profile obtained for the structure with the doping imbalance.

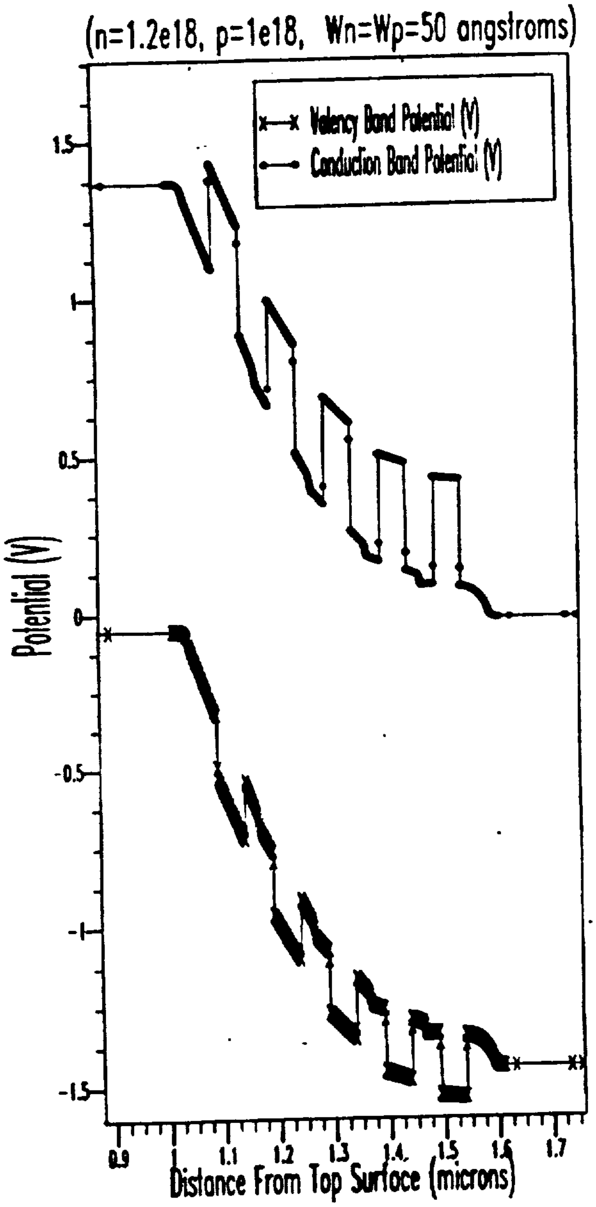

Figure 4.2.10 Band diagram of the structure with the doping imbalance showing non-depleted stages of the device.

to be able to deplete as many stages as possible which requires the need for precise doping control during MBE growth. As was previously reported, we were recently able to successfully fabricate several 10-period doped-well APDs that demonstrated full depletion below avalanche breakdown. However, when the number of built-in GaAs wells was increased to 25, the APDs displayed only partial depletion. This prompted us to try and determine the threshold limit beyond which further depletion ceases to occur. Therefore, we are currently in the process of growing and fabricating additional doped-well APDs where the number of stages in the avalanche region ranges from 10 to 25. This will help us determine the optimal width and periodicity of the depletion region which will maximize the gain at low bias in these structures.

For future work, we will extend our simulation further to gain more insights into the properties of these structures at low and high bias when impact ionization becomes important. This will eventually lead to constructing better theoretical models of the gain, noise and bandwidth characteristics of the APD which can be used to analyze our existing experimental data and enable us to use such models as a basis for fabricating high performance APD structures.

Bibliography

[2.1.1] E. G. Bogus, "Electrical Charge Injection in an Acoustic Charge Transport Device," Ph. D. Dissertation, University of Illinois, Champaign, IL, 1987.

[2.1.2] S. M. Knapp, J. J. Liou, and D. C. Malocha, "Modelling the Charge Injection Process in Acoustic Charge Transport Devices," *Proc. 1989 Ultrasonics Symp.* Dec 4-7, 1989.

[2.1.3] G. A. Peterson, B. J. McCartin, W. J. Tanski, and R. E. LaBarre, "Charge Confinement in Heterojunction Acoustic Charge Transport Devices," *Appl. Phys. Letters*, Vol. 55, No. 13, pp. 1330-1332, September 25, 1989.

[2.1.4] J. S. Kenney and W. D. Hunt, "A Physically-Based Small-Signal Circuit Model for Heterostructure Acoustic Charge Transport Devices," *IEEE Trans. Microwave Theory Techn.*, Vol. 41, No. 12, pp. 2218-26, Dec., 1993.

[2.1.5] M. J. Hoskins, and B. J. Hunsinger, "Simple Theory of Buried Channel Acoustic Charge Transport in GaAs," *Journal of Applied Physics*, Vol. 55, No. 2, pp. 413-26, January 15, 1984.

[2.1.6] R. L. Miller, C. E. Nothwick, and D. S. Bailey, *Acoustic Charge Transport Devices,* Boston: Artech House, 1992.

[2.1.7] W. J. Tanski, S. W. Merritt, R. N. Sacks, and D. E. Cullen, "Heterojunction Acoustic Charge Transport Devices on GaAs," *Appl. Phys. Lett.*, Vol. 52, No. 1, pp. 18-19, Jan. 4, 1988.

[2.1.8] S. W. Merritt, "Heterostructure Acoustic Charge Transport Device Model," *Proc. 1990 Ultrasonics Symp.*, pp. 247-252, Dec. 4-7, 1990.

[2.1.9] T. W. Grudkowski and R. N. Sacks, "HACT Structure with Reduced Surface State Effects," United States Patent 5264717, Nov. 23, 1993.

[2.1.10] A. W. Smith, J. S. Kenney, W. D. Hunt, K. F. Brennen, R. Benz, and C. J.. Summers, "Charge Confinement in a pn^-np Heterojunction Acoustic Charge Transport Device," abstract submitted to NUPAD VI, Oct. 1993.

[2.1.11] S. Datta, *Surface Acoustic Wave Devices.* Englewood Cliffs, NJ: Prentice-Hall, 1986.

[2.1.12] M. Feldmann and J. Henaff, *Surface Acoustic Waves for Signal Processing*, Boston: Artech House, 1989.

[2.1.13] H. K. Gummel, "A Self-Consistent Iterative Scheme for One-Dimensional Steady State Transistor Calculations," *IEEE Trans. Electron Dev.*, Oct., 1964, pp. 455-65.

[2.1.14] S. Tiwari, *Compound Semiconductor Device Physics*, Boston: Academic Press, Inc., 1992.

[2.1.15] B. C. Schmukler, *Charge Control in Acoustic Charge Transport Devices*, Ph.D. Thesis, University of Illinois, 1989.

[2.1.16] A. R. Hutson and D. L. White, "Elastic Wave Propagation in Piezoelectric Semiconductors," *J. Appl. Phys.*, Vol. 30, No. 40, pp. 40-47, 1962.

[2.1.17] D. L. White, "Amplification of Ultrasonic Waves in Piezoelectric Semiconductors," *J. Appl. Phys.*, Vol. 33, pp. 2547-54, 1962.

[2.1.18] F. Scheid, *Schaum's Outline of Theory and Problems of Numerical Analysis*, 2nd ed., New York: McGraw Hill, 1988.

[2.1.19] S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed., New York: John Wiley and Sons, 1981.

[2.1.20] H. Morcoc, H. Unlu, and G. Ji, *Principles and Technology of MODFET's*, vol. 1, New York: John Wiley and Sons, 1991.

[2.1.21] J. M. Golio, ed. *Microwave MESFETs and HEMTs*, Boston: Artech House, 1991.

[2.1.22] C. L. Liang, N. W. Cheung, R. N. Sato, M. Sokolich, and N.A. Doudoumopoulos, "A Diffusion Model of Subthreshold Current for GaAs MESFETs," *Solid-State Electron.*, Vol. 34, No. 2., pp. 131-38, 1991.

[2.1.23] Z. M. Li, D. J. Day, S. P. McAlister, and C. M. Hurd, "The Importance of Contact Injection in Undoped AlGaAs/GaAs Heterostructures," *IEEE Trans. Electron. Dev.*, Vol. 37, No. 3, pp. 556-61, March, 1990.

[2.1.24] G. N. Maracus, "Ohmic Contacts to GaAs," chapter 9 in *Gallium Arsenide Technology*, Vol. II, Carmel, Indiana: Howard W. Sams, 1990, pp. 373-98.

[2.1.25] R. Williams, *Modern GaAs Processing Techniques*, Boston: Artech House, 1990.

[2.1.26] *PC-1D* is available from Sandia National Laboratories, Albuquerque, NM, 87185.

[2.2.27] M. J. Hoskins, H. L. Hsiung, and F. O. Guediri, "High Speed GaAs Acoustic Charge Transport Multiplexer," *SPIE Vol. 1071, Optical Sensors and Electronic Photography*, Society of Photo-Optical and Instumentation Engineers, 1989.

[2.2.1] F.S.Hickernell, ``ZnO processing for bulk- and surface-wave devices," *Proc. IEEE Ultrason. Symp.* pp. 785-794, 1980.

[2.2.2] F.S.Hickernell, ``Piezoelectric thin-film materials," *Proc. Int. Symp. SAW Devices for Mobile Comm.*, pp. 31-38, 1992.

[2.2.3] B.J.Hunsinger, *Research to provide a theoretical determination of surface acoustic wave velocity and impedance differences between metal strips and free surface regions of metallic gratings, Technical Report,* ROME Air Development Center 1981.

[2.2.4] G.R.Adams, J.D.Jackson, and J.S.Heeks, ``Monolithic ZnO-GaAs acousto-electric devices," *Proc. IEEE Ultrason. Symp.*, pp. 109-112, 1980.

[2.2.5] T.S.Hickernell, F.M.Fliegel, and F.S.Hickernell, ``The elastic properties of thin-film silicon nitride," *Proc. IEEE Ultrason. Symp.*, pp. 445-448, 1990.

[2.2.6] T.S.Hickernell and F.S.Hickernell, ``The elastic properties of PECVD silicon oxynitride films on gallium arsenide," *Proc. IEEE Ultrason. Symp.*, pp. 51-54, 1991.

[2.2.7] Y.Kim and W.D.Hunt, ``Acoustic fields and velocities for surface-acoustic-wave propagation in multilayered structures: An extension of the Laguerre polynomial approach," *J. Appl. Phys.*, vol. 68, pp. 4993-4997, 1990.

[2.2.7] B.Hadimioglu, S.A.Elrod, D.L.Steinmetz, M.Lim, J.C.Zesch, B.T.Khuri-Yakub, E.G.Rawson, and C.F.Quate, ``Acoustic ink printing" *Proc. IEEE Ultrason. Symp.*, pp. 929-935, 1992.

[2.2.8] W.D.Hunt, Y.Kim, and F.M.Fliegel, ``A synopsis of surface acoustic wave propagation on {100}-cut <110>-propagating gallium arsenide," *J. Appl. Phys.*, vol. 69, pp. 1936-1941, 1991.

[2.2.9] J.Temmyo and S.Yoshikawa, ``Aluminum evaporation conditions for SAW interdigital transducers," *IEEE Trans. Sonics and Ultrason.*, vol. 27, pp. 219-221, 1980.

[2.2.10] V.E.Steel, W.D.Hunt, M.A.Emanuel, J.J.Coleman, and B.J.Hunsinger, ``Surface acoustic wave properties of aluminum arsenide," *J. Appl. Phys.*, vol. 66, pp. 90-96, 1989.

[2.2.11] W.D.Hunt, R.L.Miller, and B.J.Hunsinger, ``Slowness surface measurements for zero- and five- degree [100]-cut GaAs," *J. Appl. Phys.* vol. 60, pp. 3532-3538, 1986.

[2.2.12] A.J.Slobodnik, E.D.Conway, R.T.Delmonico, and eds., *Microwave Acoustics Handbook*, vol. 1A., Air Force Cambridge Research Laboratory, 1973.

[2.2.13] V.M.Bright, Y.Kim, and W.D.Hunt, ``Study of surface acoustic waves on the {110} plane of gallium arsenide," *J. Appl. Phys.*, vol. 71, pp. 597-605, 1992.

[2.2.14] Y.Kim, W.D.Hunt, Y.Liu, and C.K.Jen, ``Velocity surface measurements for ZnO films over {001}-cut GaAs," *J. Appl. Phys.*, vol. 76, pp. 1455-1461, Aug. 1, 1994.

[2.2.15] A.J.Slobodnik, ``Materials and their influence on performance," in *Acoustic Surface Waves, Topics in Applied Physics*, Vol.24, Chapter 6 (A.A.Oliner, ed.), New York: Springer-Verlag, 1978.

[2.2.16] F.S.Hickernell, ``Surface acoustic wave propagation loss in zinc oxide films," *Proc. IEEE Ultrason. Symp.*, pp. 325-328 , 1982.

[2.2.17] M.R.Melloch and R.S.Wagers, ``Propagation loss of the acoustic pseudosurface wave on (ZXt) 45° GaAs," *Appl. Phys. Lett.* vol. 43, pp. 1008-1009, 1983.

[2.2.18] T.W.Grudkowski and M.Gilden, ``Realization of temperature-compensated GaAs surface acoustic wave delay lines," *Appl. Phys. Lett.*, vol. 38, pp. 412-413 , 1981.

[2.2.19] W.D.Hunt and B.J.Hunsinger, ``A precise angular spectrum of plane waves diffraction theory for leaky wave materials," *J. Appl. Phys.*, vol. 64, pp. 1027-1032 , 1988.

[2.2.20] M.G.Cohen, ``Optical study of ultrasonic diffraction and focusing in anisotropic media," *J. Appl. Phys.*, vol. 38, pp. 3821-3828, 1967.

[2.2.21] B.A.Auld, *Acoustic Fields and Waves in Solids* , vol. 1. New York : Wiely , 1973.

[2.2.22] J.Kushibiki and N.Chubachi, ``Material characterization by line-focus-beam acoustic microscope," *IEEE Trans. Ultason. Ferroelec. Freq. Contr.* , vol. 32, pp. 189-212, 1985.

[2.2.23] A.Briggs, *Acoustic Microscopy*, New York : Oxford University Press , 1992. Chap.11.

[2.2.24] Y.C.Lee, J.O.Kim, and J.D.Achenbach, ``V(z) curves of layered anisotropic materials for the line-focus acoustic microscope," *J. Acoust. Soc. Amer.* , vol.

94 , pp. 923-930 , 1993.

[2.2.25] P.S.Cross, ``Reflective arrays for SAW resonators," *Proc. IEEE Ultrason. Symp.* , pp. 241-244 , 1975.

[2.2.26] S.Datta, *Surface Acoustic Wave Devices,* NJ : Prentice-Hall , 1986.

[2.2.27] P.V.Wright, ``Modeling and experimental measurements of the reflection properties of SAW metallic gratings," *Proc. IEEE Ultrason. Symp.*, pp. 54-63 , 1984.

[2.2.28] T.W.Grudkowski, G.K.Montress, M.Gilden, and J.F.Black, ``Integrated circuit compatible surface acoustic wave devices on gallium arsenide," *IEEE Trans. Microwave Theory and Techniq.* , vol. 29, pp. 1348-1356 , 1981.

[2.3.1] J. Henaff, M. LeContellec and C. Rudelle, " Amplification of Elastic Surface Waves", L'Echo des Recherches, Special Issue 1975.

[2.3.2] A.R Hutson, J.H. McFee, and D.L. White, "Ultrasonic Amplification in CdS", Phys. Rev. Lett, Vol. 7, no.7, 1961, pp. 237-239.

[2.3.3] S. Ludvik and C.F. Quate, "Amplification of Surface Shear-Wave Mode in GaAS", J.Appl.Phys, Vol 43, no.9, Sept 1972, pp 3619-3622.

[2.3.4] C. W. Turner, A. Shomon and R.M. White, "Reduced-Voltage Operation of a Surface -Elastic-Wave Amplifier", Electronic Letters, Vol 5, pp. 244-245, May 1969.

[2.3.5] G. Cambon, C. Roustan and M. Rouzeyre, "Segmented Surface-Wave Amplifier Exhibiting Gain at 86 MHz", Appl. Phys. Lett., Vol. 21, No. 11, Dec 1, 1972.

[3.2.1] J.D.E. Beynon and D.R. Lamb, *Charge-coupled devices and their applications*, London, McGraw-Hill, 1980.

[3.2.2] H.H. Wieder, "Dielectric/semiconductor interfaces," *J. Vac. Sci. Technol.*, vol. B11, pp.1331-1334, 1993.

[3.2.3] I. Deyhimy, R.C. Eden, J.S. Harris, Jr., "GaAs and related heterojunction charge-coupled devices," *IEEE Trans. Electron Dev.*, vol. 27, pp. 1172-1180, 1980.

[3.2.4] E.R. Fossum, J.-I. Song, D.V. Rossi, "Two-Dimensional Electron Gas Charge Coupled Devices (2DEG-CCD's),"., vol. 38, pp. 1182-1192, 1991.

[3.2.5] T.B. Stellwag, J.A. Cooper, Jr., and M.R. Melloch, "A vertically integrated GaAs bipolar dynamic RAM cell with storage times of 4.5 h at room temperature," *IEEE Electron Devices Lett.*, vol. 13, pp. 129-131, 1992.

[3.2.6] D.L. Hetherington, J.F. Klem, H.T. Weaver, "An integrated GaAs n-p-n-p thyristor/JFET memory cell exhibiting nondestructive read," *IEEE Electron Devices Lett.*, vol. 13, pp. 476 - 478, 1992.

[3.2.7] S. Tiwari and D.J. Frank, "Empirical fit to band discontinuities and barrier heights in III-V alloy systems," *Appl. Phys. Lett.*, vol. 60, pp. 630-632, 1992.

[3.2.8] T.J. Cunningham, E.R. Fossum, and S.M. Baier, "An analysis of the temperature dependence of the gate current in complementary heterojunction field-effect transistors," *IEEE Electron Devices Lett.*, vol. 13, pp. 645 - 647, 1992.

[3.2.9] J.S. Kleine, Q.-D. Qian, J.A. Cooper, Jr., and M.R. Melloch, "Electron emission from direct bandgap heterojunction capacitors," *IEEE Trans. Electron Dev.*, vol. 36, pp. 289 - 299, 1989.

[3.2.10] F. Beltram, F. Capasso, J.F. Walker, and R.J. Malik, "Memory phenomena in heterojunction structures: Evidence for suppressed thermionic emission," *Appl. Phys. Lett.*, vol. 53, pp. 376-378, 1988.

[3.2.11] F. Beltram, F. Capasso, J.F. Walker, and R.J. Malik, "Memory phenomena in heterojunction structures: Evidence for suppressed thermionic emission," *Appl. Phys. Lett.*, vol. 53, pp. 376-378, 1988.

[3.3.1] R.A. Metzger, M. Hafizi, W.E. Stanchina, T. Liu, R.G. Wilson, and L.G. McCray, "Confinement of high Be doping levels in AlInAs/GaInAs npn heterojunction bipolar transistors by low temperature molecular-beam epitaxial growth," *Appl. Phys. Lett.*, vol. 63, pp. 1360 - 1362, 1993.

[4.1.1] H. Shichijo and K. Hess, "Band structure dependent transport and impact ionization in GaAs," Phys. Rev. B, **23**, 4197 (1981).

[4.1.2] M. V. Fischetti and S. E. Laux, "Monte Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects," Phys. Rev. B, **38**, 9721 (1988).

[4.1.3] C. Jacoboni and P. Lugli, The Monte Carlo Method for Semiconductor Device Simulation, (Springer, Vienna, 1989).

[4.1.4] L. V. Keldysh, "Concerning the theory of impact ionization in semiconductors," Sov. Phys. JETP, **21**, 1135 (1965).

[4.1.5] J. Bude, K. Hess, and G. J. Iafrate, "Impact ionization in semiconductors: Effects of high electric fields and high scattering rates," Phys. Rev. B, **45**, 10958 (1992).

[4.1.6] N. Sano and A. Yoshii, "Impact ionization theory consistent with a realistic band structure of silicon," Phys. Rev B, **45**, 4171 (1992).

[4.1.7] M. Stobbe, R. Redmer, and W. Schattke, "Impact ionization rate in GaAs," Phys. Rev. B, **49**, 4494 (1994).

[4.1.8] Y. Wang and K. F. Brennan, "Semiclassical study of the wavevector dependence of the interband impact ionization rate in bulk silicon," J. Appl. Phys., **75**, 313 (1994).

[4.1.9] Y. Wang and K. F. Brennan, "Numerical study of the wave-vector dependence of the electron interband impact ionization rate in bulk GaAs," J. Appl. Phys., **76**, 974 (1994).

[4.1.10] J. Kolnik, Y. Wang, I. H. Oguzman, and K. F. Brennan, "Theoretical investigation of wave-vector dependent analytical and numerical formulations of the interband impact ionization transition rate for electrons in bulk silicon and GaAs," J. Appl. Phys., **76**, 3542 (1994).

[4.1.11] W. Quade, E. Schoell, and M. Rudan, "Impact ionization within the hydrodynamic approach to semiconductor transport," Solid-State Electron., **36**, 1493 (1993).

Numerical study of the wave-vector dependence of the electron interband impact ionization rate in bulk GaAs

Yang Wang and Kevin F. Brennan
School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0250

(Received 16 December 1993; accepted for publication 4 April 1994)

Ensemble Monte Carlo calculations of the electron interband impact ionization rate in bulk GaAs are presented using a wave-vector (**k**)-dependent formulation of the ionization transition rate. The transition rate is evaluated through use of numerically generated wavefunctions determined via a **k**·**p** calculation within the first two conduction bands at numerous points within a finely spaced three-dimensional grid in **k** space. The transition rate is determined to be greatest for states within the second conduction band. It is found that the interband impact ionization transition rate in bulk GaAs is best characterized as having an exceedingly "soft" threshold energy. As a consequence, the dead space, defined as the distance over which the ionization probability for a given carrier is assumed to be zero, is estimated to be much larger than that estimated using a "harder" threshold. These results have importance in the design of multiquantum-well avalanche photodiodes.

I. INTRODUCTION

The probability that an interband impact ionization event will occur depends upon the likelihood that a carrier will first drift to an energy equal to or greater than the band-gap energy, often referred to as the threshold energy, and then suffer an impact ionization event before phonon scattering processes relax it back to an energy below threshold. Most of the early analytical theories of interband impact ionization[1,2] concerned themselves with determining the probability that a carrier survives to threshold assuming that it would ionize immediately afterward. Such theories are said to assume a "hard threshold" for interband impact ionization.

The hard threshold model was not limited to the early analytical theories. More advanced theories[3,4] were formulated using a hard threshold model for interband impact ionization; however, more recent numerical studies[5,6] as well as analytical theories[7,8] have called this assumption into question. Interestingly, in most of the numerical theories employed in the past the calculated impact ionization rates often agree well with experimental results for some semiconductor materials using either a hard or "soft" threshold model.[9] This is due to the fact that most numerical formulations rely on a parameterized expression, called the Keldysh formula,[10] to describe the impact ionization transition rate. The Keldysh formula contains two parameters, P and E_{th}, which are very difficult to determine from first principles. E_{th} corresponds to the threshold energy for impact ionization while P is some measure of how strong the impact ionization transition rate is for carriers with energies greater than E_{th}. A small value of P then corresponds to a soft threshold since the transition rate is relatively small near the threshold and increases only slowly thereafter. In contrast, a hard threshold is characterized by a large value of P. Typically, both P and E_{th} are selected such that the calculated impact ionization rate is in agreement with experimental measurements. As such, more than one set of parameters can often reproduce the same ionization rate in a given material. Tang and Hess[9] have shown that the calculated impact ionization rate in bulk silicon, based on an ensemble Monte Carlo model including the Keldysh formula, agrees well with existing experimental data using either a soft or hard threshold model.

Kane[11] made the first attempt to improve the impact ionization transition rate formulation by directly evaluating the pair cross section using pseudopotential-generated energy bands and wave functions for bulk silicon. In his formulation, Kane calculated the transition rate for different wave vector (**k** vector) states of comparable energy and averaged these rates to determine an energy dependent transition rate. More recently, Sano *et al.*[12,13] and Bude and Hess[5] have investigated the threshold energy in silicon and GaAs using numerical techniques. Their work indicates that the threshold is highly **k** dependent and that the relative softness of the threshold originates from the **k** dependence of the impact ionization transition rate.

New theories of interband impact ionization have been presented which treat the problem either semiclassically [5,14] or quantum mechanically.[15–17] The quantum-mechanical works have demonstrated that the threshold energy is always relatively soft owing to energy broadening effects arising from the electron-phonon interaction. In the quantum-mechanical formulations, a Lorentzian is substituted in place of the energy conserving delta function which leads to a broadening of the threshold state itself. Bude and co-workers[15] have demonstrated that through inclusion of quantum-mechanical broadening, a threshold state cannot rigorously be defined since the Lorentzian does not demand strict energy conservation over short time scales. As a result, a carrier can impact ionize even if its energy is less than the energy gap. Energy broadening acts then to inflate the impact ionization rate near threshold.

Similarly, even the semiclassical models predict that the ionization transition rate in bulk silicon is relatively low leading to a soft thresholdlike behavior. Owing to the fact that the ionization transition rate is calculated to be significantly lower than the competing phonon scattering rate, these models predict that the electrons drift to high energies, on average, before suffering an impact ionization event. As a

0021-8979/94/76(2)/974/8/$6.00

© 1994 American Institute of Physics

result, the average energy of the electrons is expected to be quite high at high-electric-field strengths in contrast to what is predicted using a hard threshold model. Recent experimental measurements using x-ray photoemission spectroscopy (XPS)[18] indicate that the threshold for interband impact ionization in bulk silicon is exceedingly soft. These results further indicate that the ionization rate may have multiple ranges of behavior which is consistent with theoretical predictions[5,14] that the second band dominates the impact ionization process in bulk silicon.

Aside from the obvious importance of improving the understanding of interband impact ionization, knowledge of the nature of the threshold energy in different materials is useful in modeling advanced avalanche photodiode (APD) designs which utilize heterostructures and multiple quantum wells to enhance the electron ionization rate.[19–22] In these devices the presence of the conduction-band-edge discontinuity formed at the heterostructure interface provides a kinetic energy boost to electrons injected into the narrow-band-gap layers from the wide-band-gap materials. The larger the conduction-band-edge discontinuity is, the greater relative effect it has on the electron ionization rate. This depends though upon how soft the threshold is. If the threshold is relatively hard, the conduction-band-edge discontinuity can greatly influence the ionization rate by providing sufficient energy to promote the electron from below threshold to above threshold. On the other hand, if the threshold is exceedingly soft, the presence of the conduction-band-edge discontinuity may not have as much of an effect since most of the carriers must attain very high energies before impact ionizing. As a result, the discontinuity may contribute relatively little to the overall heating of the electron distribution.

Since interband impact ionization is a threshold process, a carrier launched near the band edge, either initially or after suffering an ionization event, must drift for some time to acquire sufficient kinetic energy in order to impact ionize. The ionization probability then is necessarily equal to zero for some distance for both the generated carrier and the initiating carrier immediately following an ionization event. The distance over which the ionization probability is assumed to be zero following launching from the band edge is often referred to as the dead space. The dead space can be of importance in multiquantum-well structures since if the well widths are much smaller than the dead space distance, then the ionization rate may be appreciably reduced, leading to little enhancement. Recently, the electron and hole ionization coefficients were measured in a series of GaAs/AlGaAs multiple quantum wells with varying well and barrier widths.[23] It was found that in the narrow well width structures (<100 Å) there was virtually no enhancement of the experimentally measured electron to hole ionization rates ratio, while in wider well structures there was a significant increase in the ionization rates ratio. These results may be related to the dead space in the ionization process.

In this article we present detailed calculations of the electron-initiated impact ionization rate in bulk GaAs using a semiclassical formulation of the interband impact ionization transition rate. The transition rate is determined using Fermi's golden rule from a two-body screened Coulomb interaction assuming both energy and momentum conservation for the first two conduction bands of GaAs. These results are then incorporated into an ensemble Monte Carlo simulator to calculate the total impact ionization rate as a function of applied electric field in bulk GaAs. Aside from the ionization transition rate and the total impact ionization rate, the dead space for electron-initiated impact ionization is also calculated and compared to results based on a relatively hard threshold model. The implications of an exceedingly soft threshold for impact ionization in GaAs are discussed in relation to multiquantum-well devices. The details of the model are summarized in Sec. II. The calculated results are presented in Sec. III while conclusions are drawn in Sec. IV.

II. MODEL DESCRIPTION

The details of the model have been exhaustively reviewed in a previous work.[14] Nevertheless, for completeness, the salient features of our model are briefly outlined here. The total impact ionization rate is determined using an ensemble Monte Carlo simulation which includes the full details of the first two conduction bands of GaAs and the numerically generated semiclassical impact ionization transition rate. The Monte Carlo simulator used in the calculations is based on the original work of Shichijo and Hess,[24] with the improved phonon scattering technique of Chang *et al.*[25] and the improved band-structure interpolation technique of Fischetti and Laux.[26] The material parameters used in the simulation are taken from Refs. 27 and 28. The estimators used in the simulation are discussed at length in Ref. 28.

The high-energy scattering rate is assumed to be dominated by deformation potential scattering. The deformation potential scattering rate is calculated using a full order time-dependent perturbation theory expansion assuming that the deformation potential is a constant, independent of energy.[25] Although the assumption of a constant deformation potential may not be strictly valid,[17] it is standard and commonly used. The deformation potential scattering is determined by direct integration over the numerically generated density of states for both bands. At the points in **k** space at which both bands overlap, the deformation potential scattering depends, of course, on the density of states within both bands. During the course of the Monte Carlo simulation, the electrons can transfer then from one band to the other through the action of the deformation potential scattering.

Impact ionization is incorporated into the Monte Carlo model using the semiclassical, wave-vector (**k**)-dependent transition rate given by Ridley.[29] The matrix element is assumed to be a screened Coulomb interaction between two electrons.[14] The transition rate is evaluated numerically by integrating over the full Brillouin zone. Several million possible final states are sampled for each integration to ensure convergence and accuracy. The overlap integrals present in the matrix element are determined by diagonalizing a 15×15 $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian.[30] The values of the overlap integrals match closely to those presented by Burt *et al.*[31] The transition rate is determined for each mesh point within a finely spaced, 916 point, **k**-space grid spanning the reduced zone[24] of the first Brillouin zone. Additional points are also used for interpolation purposes raising the total number of mesh points at

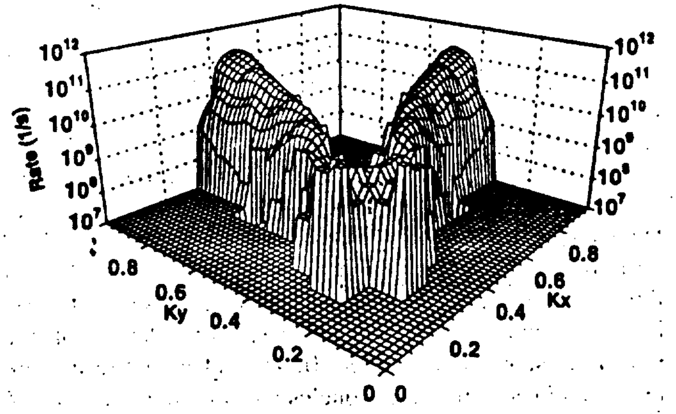

FIG. 1. Calculated interband impact ionization transition rate for the first conduction band of bulk GaAs within the reduced zone of the first Brillouin zone. The ionization transition rate is plotted vs k_x and k_y holding k_z fixed at 0.0. Those points which are shown to have a transition rate equal to the minimum of the vertical axis scale have no threshold and hence have a zero transition rate. Owing to the logarithmic scale used in the figure, these points are represented as having a minimum transition rate for convenience.

which the transition rate is calculated to 1419. The transition rate is determined for both conduction bands at each of the mesh points. The mesh employed for the impact ionization transition rate is the same as that for the energy interpolation used during the course of the Monte Carlo simulation.[26]

The instantaneous electron-impact ionization transition rate is determined both from its energy and **k** vector. When the simulated electron's energy is less than the energy band gap, the electron cannot impact ionize and its flight is determined in the usual manner.[32] When the simulated electron drifts to an energy greater than or equal to the energy band gap, its ionization transition rate is determined using the **k** space mesh. The eight mesh points nearest to the components of the simulated electron's **k** vector are first determined. If the ionization transition rate is nonzero for all of the eight mesh points then the transition rate of the simulated electron is determined by linearly interpolating between the values at each mesh point. If one or more of the mesh point values is zero, then the transition rate for the simulated electron is chosen equal to that of the nearest **k** point. The impact ionization transition rate determined in this manner is then added to the total phonon scattering rate corresponding to the energy of the simulated electron. As in standard Monte Carlo algorithms, a random number is thrown to determine if the carrier is scattered. Given that the carrier scatters, a second random number is generated to determine the particular event that occurs out of all the possible events at that energy and **k** vector.

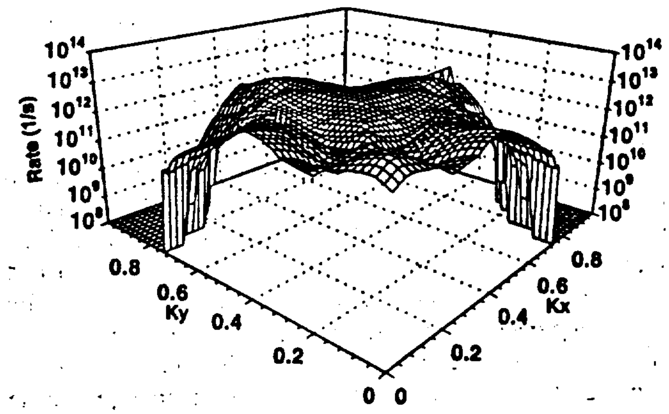

FIG. 2. Calculated interband impact ionization transition rate for the second conduction band of bulk GaAs within the reduced zone of the first Brillouin zone. The ionization transition rate is plotted vs k_x and k_y holding k_z fixed at 0.0. Those points which are shown to have a transition rate equal to the minimum of the vertical axis scale have no threshold and hence have a zero transition rate. Owing to the logarithmic scale used in the figure, these points are represented as having a minimum transition rate for convenience.

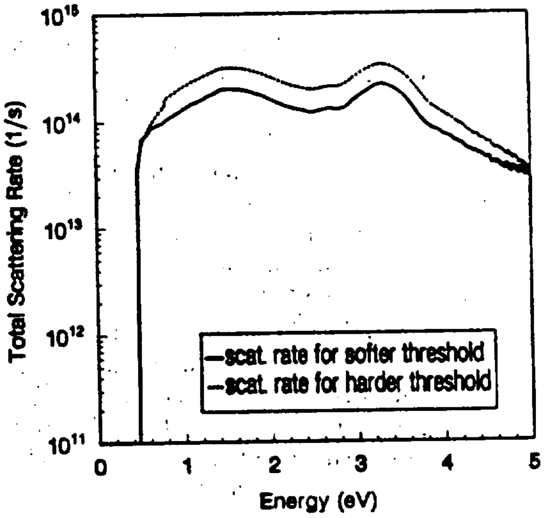

FIG. 3. Total calculated electron phonon scattering rate used in the soft threshold and hard threshold simulations. The high-energy scattering rate for the soft threshold case is matched at 0.5 eV with the phonon scattering rate calculated from Fermi's golden rule, while the rate for the hard threshold case is matched at 0.8 eV.

The calculated interband impact ionization transition rate for the first two conduction bands of bulk GaAs as a function of k_x and k_y holding k_z fixed at 0.0 within the reduced zone is presented in Figs. 1 and 2. As can be seen from comparison of these two figures, the transition rate is significantly greater in the second conduction band than in the first. The maximum transition rate in either band is only slightly greater than 10^{13} s^{-1}, significantly less than the competing phonon scattering rate at these energies. The probability that an ionization event will occur at any given time, relative to a competing phonon scattering event, is then quite small. As a result, an electron can drift for some time to very high energy before suffering an impact ionization event.

III. CALCULATED RESULTS

At high carrier energies the phonon scattering rate is assumed, as stated above, to be dominated by deformation potential scattering. The scattering rate at these energies is determined by directly integrating over the final density states calculated from the numerically generated band structure including collision broadening. A single deformation potential constant is selected in order to match the scattering

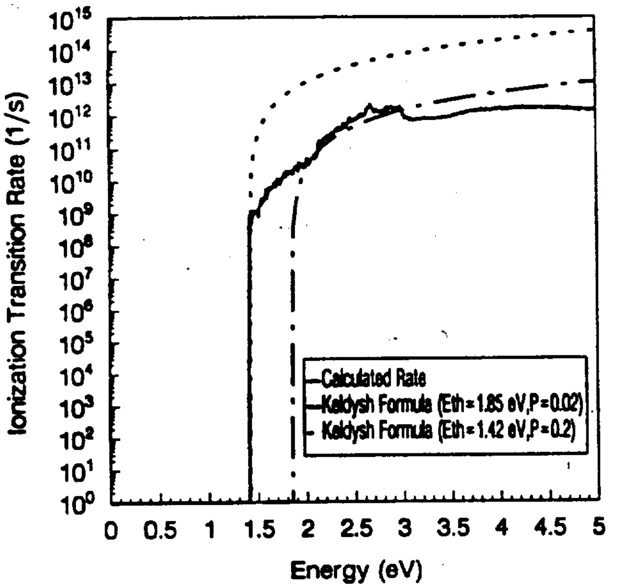

FIG. 4. Calculated impact ionization transition rates as a function of electron energy measured from the conduction-band minimum using the k-dependent and Keldysh formula with two different sets of P and E_{th}. The energy-dependent rate for the k-dependent rate is calculated by performing a further integration of the k-dependent rate.

rate calculated from Fermi's golden rule at a specific energy. Since the impact ionization coefficients are very sensitive to the total scattering rate, a small change in the total scattering rate will lead to a dramatic change in the impact ionization coefficients. Figure 3 shows the total electron phonon scattering rates with the deformation potential scattering rate matched to the scattering rate calculated from Fermi's golden rule at 0.5 and 0.8 eV, respectively. As is discussed later, the lower scattering rate, that matched at 0.5 eV, results in a softer threshold for the impact ionization process, while the higher scattering rate, that matched at 0.8 eV, results in a harder threshold. It should be noted that there is some arbi-

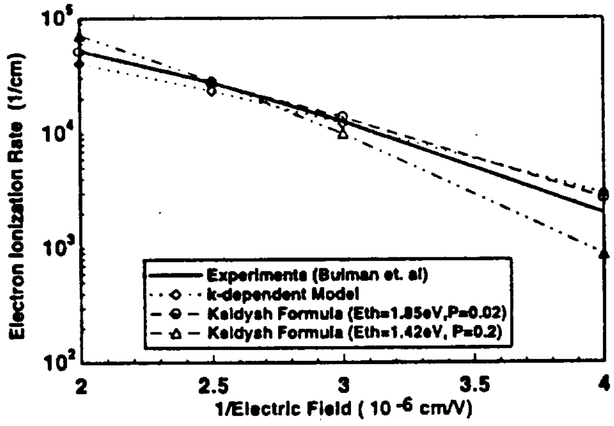

FIG. 5. Calculated and experimental results for the electron-impact ionization coefficients in bulk GaAs plotted as a function of inverse electric field, using the k-dependent model, the Keldysh formula model with $P=0.02$ and $E_{th}=1.85$ eV, and the Keldysh formula model with $P=0.2$ and $E_{th}=1.42$ eV. The experimental results are taken from Bulman *et al.* (Ref. 33).

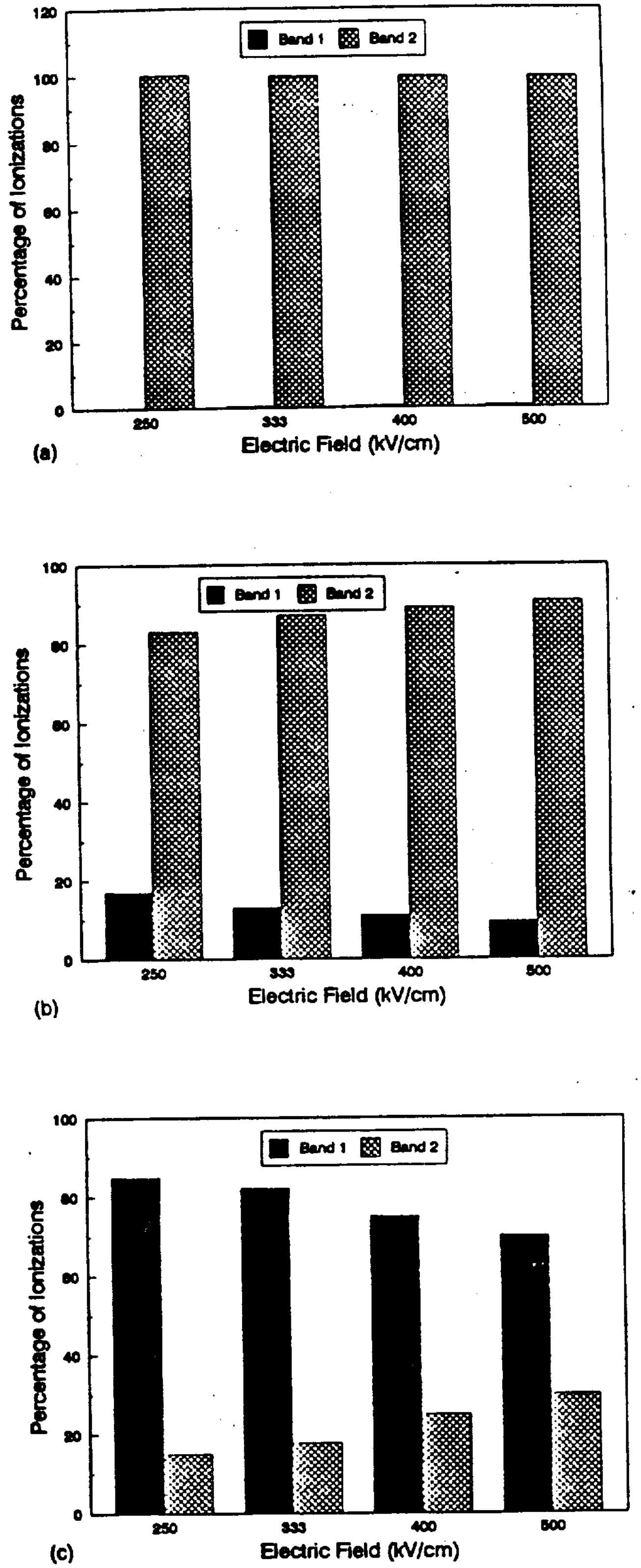

FIG. 6. (a) Histogram showing the percentage of electron ionization events originating from the first and second conduction bands as a function of applied electric field based on the k-dependent model. (b) Histogram showing the percentage of electron ionization events originating from the first and second conduction bands as a function of applied electric field based on the Keldysh formula model with $P=0.02$, $E_{th}=1.85$ eV. (c) Histogram showing the percentage of electron ionization events originating from the first and second conduction bands as a function of applied electric field based on the Keldysh formula model with $P=0.2$, $E_{th}=1.42$ eV.

trariness in the total scattering rate since the deformation potentials are not fully known as a function of energy. Typically, the scattering rate is adjusted such that the calculated carrier drift velocities agree with the experimental measurements; however, the drift velocity is less sensitive to the scattering rate than the impact ionization coefficients. In these calculations, both sets of scattering rates show good agreement with the experimental saturation carrier drift velocity within the error tolerance of the Monte Carlo technique. Therefore, either total scattering rate is plausible.

An energy-dependent impact ionization transition rate can also be derived from the **k**-dependent rate by performing an additional integration over the initiating electron **k**-vector space, normalized by the density of states. The result is plotted as a function of initiating electron energy in Fig. 4, along with the rates using the Keldysh formula with two sets of P and E_{th}; $P=0.02$ and $E_{\text{th}}=1.85$ eV, as well as $P=0.2$ and $E_{\text{th}}=1.42$ eV. Multiple sets of P and E_{th} may be found by fitting the calculated impact ionization coefficients to the experiment. The first set of values of $P=0.02$ and $E_{\text{th}}=1.85$ eV is chosen so as to match the Keldysh rate as close as possible to the energy-dependent rate derived from the **k**-dependent rate. The value of 0.02 for P is in the range of values which typically corresponds to a soft threshold condition. Another set of parameters for the Keldysh formula, $P=0.2$ and $E_{\text{th}}=1.42$ eV, is chosen to compare the difference between the hard and soft threshold models.

The impact ionization coefficients calculated using the **k**-dependent impact ionization transition rate, the Keldysh formula with $P=0.02$ and $E_{\text{th}}=1.85$ eV, and the Keldysh formula with $P=0.2$ and $E_{\text{th}}=1.42$ eV along with the experimental measurements of Bulman *et al.*[33] are plotted in Fig. 5 at an electric-field direction of [100]. The lower total scattering rate as shown in Fig. 3 is used in the calculations of both the **k**-dependent model and the Keldysh formula with $P=0.02$ and $E_{\text{th}}=1.85$ eV. In the case of the Keldysh formula with $P=0.2$ and $E_{\text{th}}=1.42$ eV the higher total scattering rate must be used due to the higher impact ionization transition rate and the harder threshold present in this case in order to agree with the experimental results. As can be seen from Fig. 5, the **k**-dependent calculations match the experiment fairly well throughout the range of applied electric-field strengths plotted here. This is excellent considering there is no adjustable parameter used in the model for the impact ionization transition rate. The only ambiguity in this case appears in the scattering rate. The calculated results using the Keldysh formula for the impact ionization transition rate, within the otherwise identical Monte Carlo model, are included for comparison. With two adjustable parameters, P and E_{th} it is possible to have better agreement with the experiments than the **k**-dependent model, as shown here for the case of $P=0.02$ and $E_{\text{th}}=1.85$ eV. The results calculated using the **k**-dependent model are closer to the results calculated by the Keldysh formula with a softer threshold as expected, since the **k**-dependent model also has effectively a very soft threshold.

The effect of the second conduction band on the impact ionization coefficients in bulk GaAs for all three impact ionization transition rate models has also been investigated. Figure 6(a) shows the percentage of electron ionization events originating from the first and second conduction bands within the **k**-dependent model. Nearly 100% of the impact ionization events originate from the second conduction band due to the fact that the impact ionization transition rate is much higher in the second conduction band. The percentage of electron ionization events originating from the first and second conduction bands for both Keldysh formula models are plotted in Figs. 6(b) and 6(c). In contrast to the **k**-dependent model, more ionization events originated from the first conduction band in both cases. This can be understood since in the Keldysh formula, the impact ionization transition rate is isotropic and no distinction is made between states in the first and second conduction bands. Therefore, the electrons have the same ionization probability at a given energy from the first conduction band as from the second conduction band in the Keldysh formula model, while in the **k**-dependent model the ionization probability is typically higher in the second conduction band than the first conduction band even at the same energy. Hence, it is not surprising that the Keldysh formula model predicts that more events will occur from the first conduction band than the **k**-dependent model does. As the threshold becomes lower and harder, more ionization events occur in the first conduction band. Therefore, the Keldysh formula model will predict very different physical results if a different set of P and E_{th} is used.

It is interesting to examine the impact ionization coefficients along different applied electric-field directions. The directional dependence of the impact ionization transition rate is incorporated naturally in the **k**-dependent formulation. Calculations of the electron-impact ionization coefficient at different applied electric field directions of [100], [110], and [111] show no field directional dependence of the impact ionization coefficients even with the **k**-dependent impact ionization transition rate. These results agree with previous calculations by Brennan and Hess[34] using the Keldysh formula and the calculations reported here using both Keldysh formula models. As shown in Fig. 6(a), almost all the impact ionization events originate from the second conduction band using the **k**-dependent formulation. Since the impact ionization transition rate in the second conduction band, shown in Fig. 2, is higher and more uniform than the rate in the first conduction band, little electric-field directional dependence of the electron-impact ionization coefficients is expected.

The electron number density functions, defined here as the product of the electron density of states function and the electron distribution function, are plotted in Figs. 7(a)–7(c) for all three models at an applied electric field of 250 kV/cm. The number density functions at the applied electric field of 500 kV/cm for all three models are presented in Figures 8(a)–8(c). Both the **k**-dependent model and the Keldysh formula model with the softer threshold predict similar electron distributions. The second peak in the distributions at an electric field of 500 kV/cm for the **k**-dependent model and the Keldysh formula with softer threshold indicate many electrons occupy the second conduction band at high applied electric fields. In the Keldysh formula model with the harder threshold, the distribution is much cooler. In this case, fewer

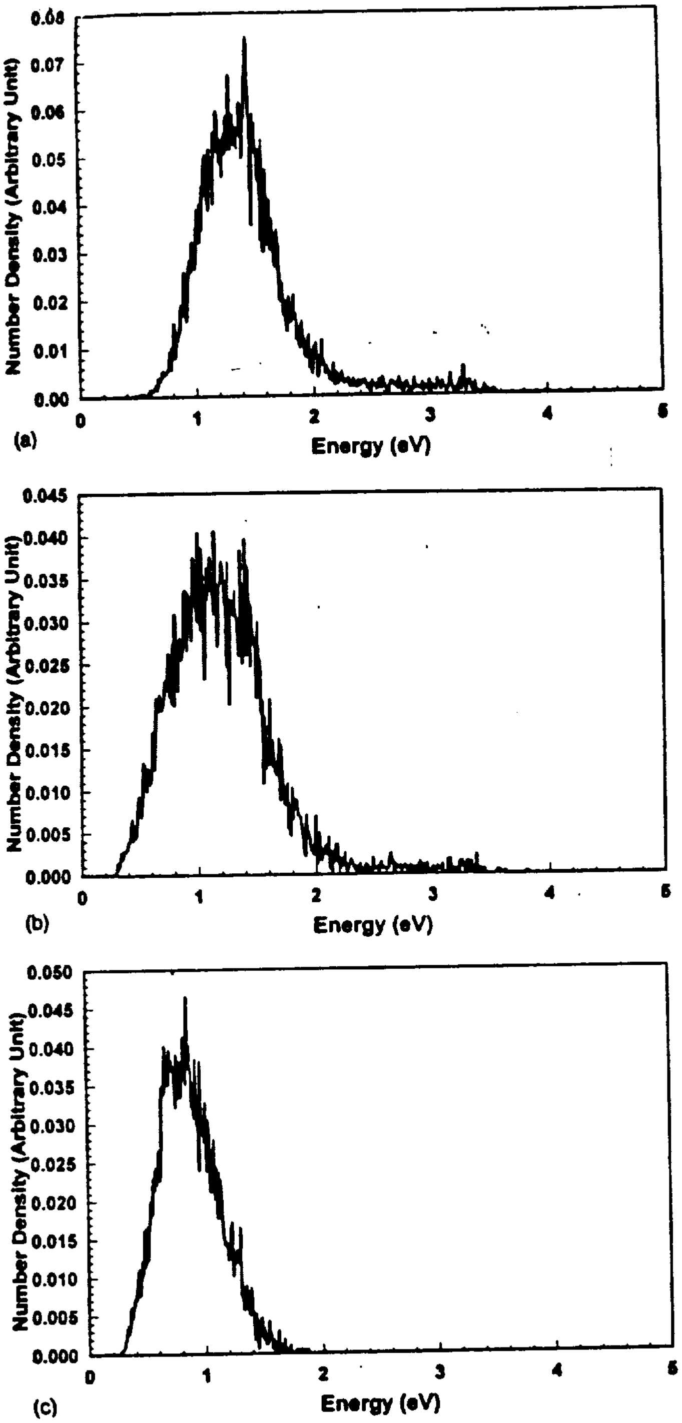

FIG. 7. (a) The electron number function, defined as the product of the electron density-of-states function and the electron distribution function, at the applied electric field of 250 kV/cm calculated using the k-dependent model. Notice the significant number of electrons within the high-energy tail of the distribution. In the k-dependent model owing to the ultrasoft threshold present, the impact ionization process does not truncate the high-energy tail of the distribution. (b) The electron number function, defined as the product of the electron density-of-states function and the electron distribution function, at the applied electric field of 250 kV/cm calculated using the Keldysh formula model with $P=0.02$, $E_{th}=1.85$ eV. As in the k-dependent model, the high-energy tail of the electron distribution is not truncated due to the soft threshold in the ionization process. (c) The electron number function, defined as the product of the electron density-of-states function and the electron distribution function, at the applied electric field of 250 kV/cm calculated using the Keldysh formula model with $P=0.2$, $E_{th}=1.42$ eV. Notice that few electrons exist within the high-energy tail of the distribution in this case. The presence of a hard threshold acts to sharply truncate the high-energy tail at relatively low carrier energy.

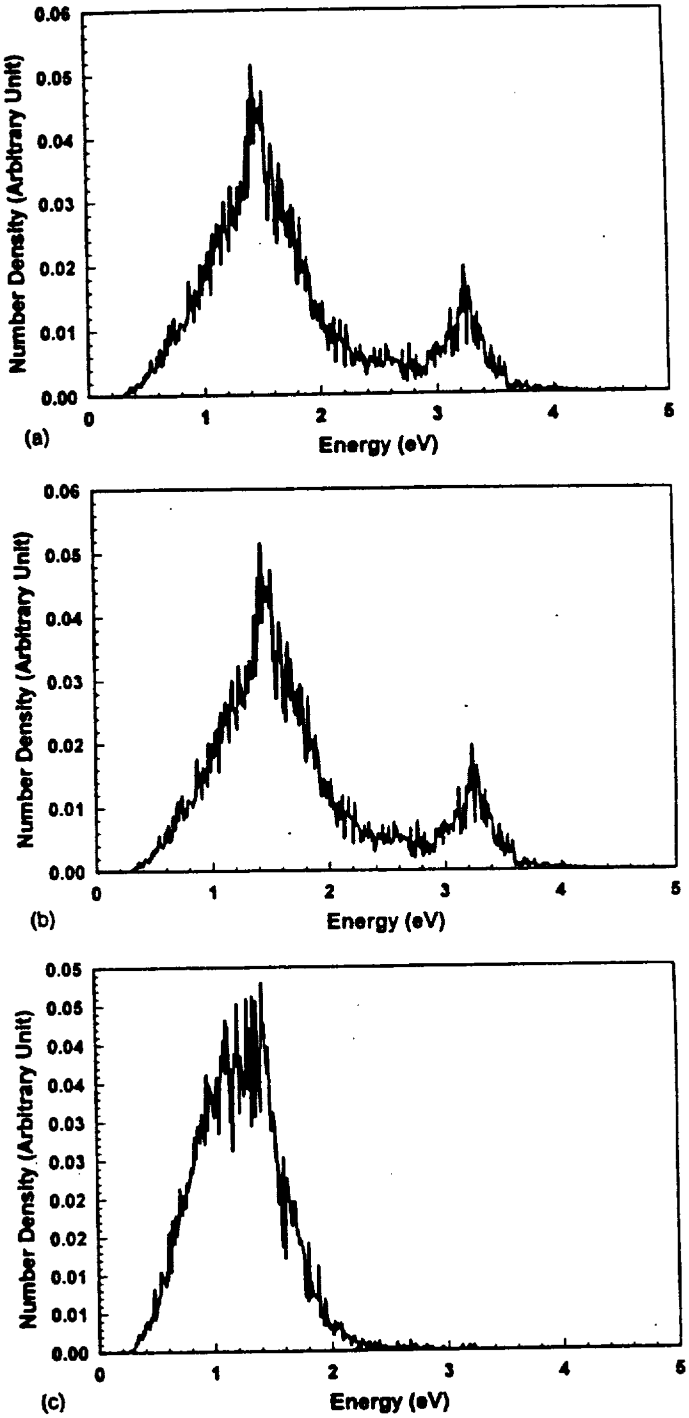

FIG. 8. (a) The electron number function, defined as the product of the electron density-of-states function and the electron distribution function, at the applied electric field of 500 kV/cm calculated using the k-dependent model. The second peak in the distribution is due to carriers within the second conduction band. Notice that there is a significant population of electrons at very high energy due to the presence of a very soft threshold for impact ionization. (b) The electron number function, defined as the product of the electron density-of-states function and the electron distribution function, at the applied electric field of 500 kV/cm calculated using the Keldysh formula model with $P=0.02$, $E_{th}=1.85$ eV. As in the k-dependent case, the second peak in the distribution is due to carriers within the second conduction band. Many electrons survive to high energy in this case owing to the very soft threshold for impact ionization. (c) The electron number function, defined as the product of the electron density-of-states function and the electron distribution function, at the applied electric field of 500 kV/cm calculated using the Keldysh formula model with $P=0.2$, $E_{th}=1.42$ eV. Owing to the relatively hard threshold used in this case, the high-energy tail of the distribution is effectively truncated.

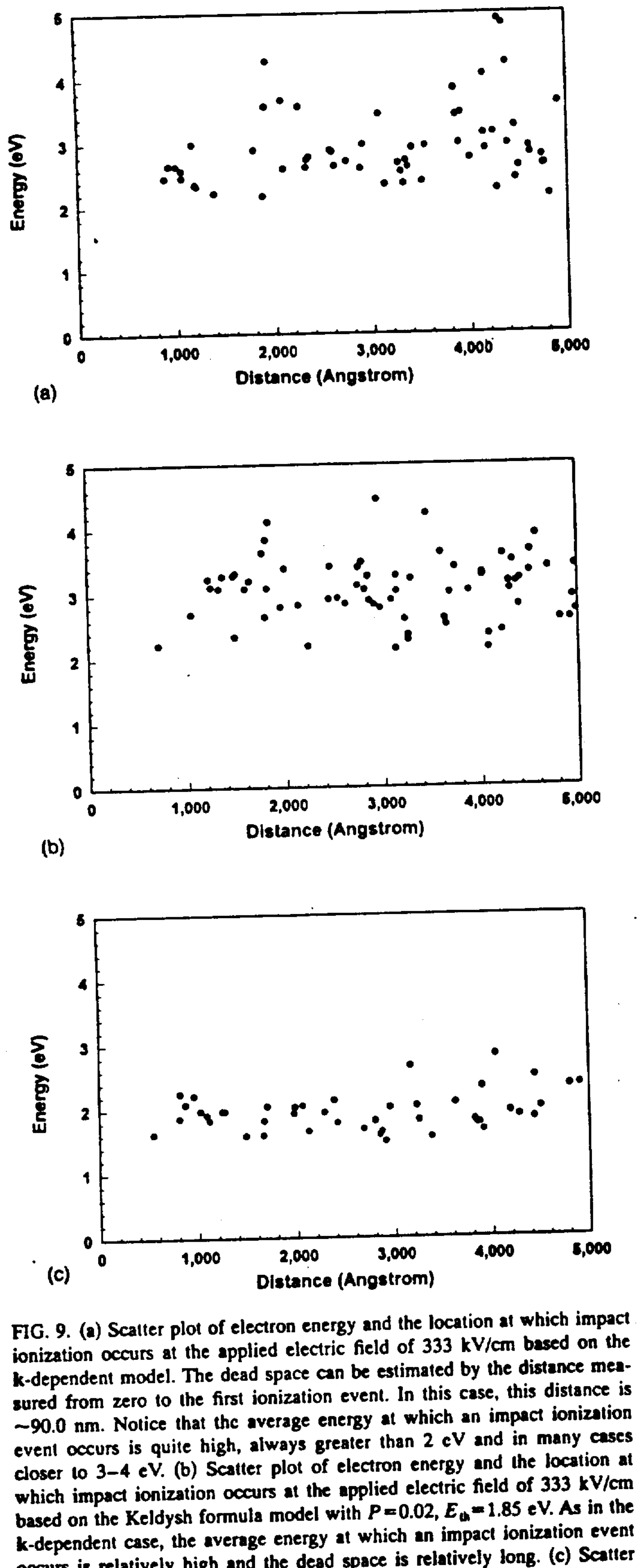

FIG. 9. (a) Scatter plot of electron energy and the location at which impact ionization occurs at the applied electric field of 333 kV/cm based on the k-dependent model. The dead space can be estimated by the distance measured from zero to the first ionization event. In this case, this distance is ~90.0 nm. Notice that the average energy at which an impact ionization event occurs is quite high, always greater than 2 eV and in many cases closer to 3–4 eV. (b) Scatter plot of electron energy and the location at which impact ionization occurs at the applied electric field of 333 kV/cm based on the Keldysh formula model with $P=0.02$, $E_{th}=1.85$ eV. As in the k-dependent case, the average energy at which an impact ionization event occurs is relatively high and the dead space is relatively long. (c) Scatter plot of electron energy and the location at which impact ionization occurs at the applied electric field of 333 kV/cm based on the Keldysh formula model with $P=0.2$, $E_{th}=1.42$ eV. Notice that the average energy at which an impact ionization event occurs is relatively low. Many ionizations occur at energies less than 2 eV and few carriers survive to energies much greater than 2 eV. Notice also that the dead space is predicted to be much less than in the k-dependent model. In this case, the dead space is ~50.0 nm almost half of that predicted for the k-dependent model.

electrons survive to high energies. This is, of course, due to the fact that in the hard threshold model, once a carrier attains threshold it is highly likely to suffer an impact ionization event. As a result, few carriers drift to energies significantly higher than the threshold.

The electron energies and positions at which impact ionizations occur are plotted in Figs. 9(a)–9(c) for all three models at the applied electric field of 333 kV/cm. Each dot in the figures represents an impact ionization event that occured at energy E and distance D from launch. The electrons are all launched initially from zero energy. From inspection of the figures, the electron-impact ionization dead space can be estimated at each specific applied electric field. Notice that the dead space predicted by the harder threshold model is the shortest among the three models. This is because in the harder threshold model electrons can impact ionize at much lower energies [Fig. 9(c)] due to the lower threshold and the higher impact ionization probability once the carrier reaches the threshold. The distance traveled for the electrons to reach the ionization energy thus is much shorter. The k-dependent model has the longest dead space. In the k-dependent model most impact ionizations occur in the second conduction band, while in the Keldysh formula model electrons can more easily impact ionize within the first conduction band. Since electrons can only transfer to the second conduction band through scattering events in our simulation model, it takes a longer time and, thus, a longer distance, for the electrons to impact ionize. Due to the use of a higher and softer threshold, electrons need reach higher energy to impact ionize. This results in longer dead spaces than that reported in Ref. 35.

IV. CONCLUSIONS

In this article we have presented an analysis of the interband impact ionization rate in bulk GaAs using a semiclassical, k-dependent, numerical formulation of the impact ionization transition rate. It is found that the transition rate is substantially higher within the second conduction band than the first conduction band. As a result, the overwhelming majority of electrons impact ionize from states within the second conduction band than from the first conduction band, in contrast to what is predicted by a hard threshold, Keldysh formula model. In addition, the numerical, k-dependent formulation of the transition rate predicts that the electrons survive to very high energy, on average, before suffering an impact ionization event. Again, this is vastly different from that predicted by the hard threshold, Keldysh formula model wherein most of the electrons ionize at much lower energies.

Interestingly, both the hard threshold model, and the soft threshold models, i.e., either a soft threshold, Keldysh formula model or the numerical, k-dependent model, can be fashioned, by adjusting the total phonon scattering rate, to match the experimental results exceedingly well. Though the hard and soft threshold models require different total phonon scattering rates in their description, the two different rates examined here both lead to acceptable predictions of the steady-state electron drift velocities. As such, it is difficult under present circumstances to definitively choose which scattering rate is the most appropriate. A more detailed analysis of the high-energy scattering rate, taking into account the energy dependence of the deformation potentials, needs ultimately to be done. Nevertheless, it is expected that the **k**-dependent transition rate formulation is far more accurate than the Keldysh formula. In order that the **k**-dependent transition rate calculations agree with the experimental measurements, the phonon scattering rate would have to be lower than that usually quoted for a hard threshold model.

If the threshold for interband impact ionization is indeed very soft in GaAs, as indicated by the present **k**-dependent transition rate calculations, the dead space for impact ionization is then significantly larger than previously estimated. A longer dead space may lead to different predictions of the electron-impact ionization enhancement in GaAs/AlGaAs multiquantum-well (MQW) structures. As a result, the optimal well and barrier widths of these devices may be quite different from those predicted using a harder threshold model. We are currently investigating how an ultrasoft threshold, as predicted by the **k**-dependent transition rate calculations presented here, would effect the predicted electron and hole ionization rates in MQW structures. These results will be presented elsewhere.

ACKNOWLEDGMENTS

This work was supported in part by ARPA and NASA through contract to NASA, Contract No. NAGW-2753. Additional support was received by the National Science Foundation through a Presidential Young Investigator Award made to K.B. and from Bell Northern Research through an industrial matching PYI award, Grant No. 125A3J. Computing support was provided by the Digital Equipment Corp. through Contract No. E21-H36.

[1] W. Shockley, Solid-State Electron. **2**, 35 (1961).
[2] P. A. Wolff, Phys. Rev. **95**, 1415 (1954).
[3] H. Shichijo and K. Hess, Phys. Rev. B **23**, 4197 (1981).
[4] B. K. Ridley, J. Phys. C Solid-State Phys. **16**, 3373 (1983).
[5] J. Bude and K. Hess, J. Appl. Phys. **72**, 3554 (1992).
[6] R. Thoma, H. J. Peifer, W. L. Engl, W. Quade, R. Brunetti, and C. Jacoboni, J. Appl. Phys. **69**, 2300 (1991).
[7] J. S. Marsland, Solid-State Electron. **30**, 125 (1987).
[8] R. C. Woods, Appl. Phys. Lett. **52**, 65 (1988).
[9] J. Y. Tang and K. Hess, J. Appl. Phys. **54**, 5139 (1983).
[10] L. V. Keldysh, Zh. Eksp. Teor. Fiz. **48**, 1962 (1965) [Sov. Phys. JETP **21**, 1135 (1965)].
[11] E. O. Kane, Phys. Rev. **159**, 624 (1967).
[12] N. Sano, T. Aoki, and A. Yoshii, Appl. Phys. Lett. **55**, 1418 (1989).
[13] N. Sano, M. Tomizawa, and A. Yoshii, Appl. Phys. Lett. **56**, 653 (1990).
[14] Y. Wang and K. F. Brennan, J. Appl. Phys. **75**, 313 (1994).
[15] J. Bude, K. Hess, and G. J. Iafrate, Phys. Rev. B **45**, 10 958 (1992).
[16] W. Quade, F. Rossi, and C. Jacoboni, Semicond. Sci. Technol. **7**, B502 (1992).
[17] P. D. Yoder, J. M. Higman, J. Bude, and K. Hess, Semicond. Sci. Technol. **7**, B357 (1992).
[18] E. Cartier, M. V. Fischetti, E. A. Eklund, and F. R. McFeely, Appl. Phys. Lett. **62**, 3339 (1993).
[19] R. Chin, N. Holonyak, Jr., G. E. Stillman, J. Y. Tang, and K. Hess, Electron Lett. **16**, 467 (1980).
[20] F. Capasso, in *Semiconductors and Semimetals*, edited by R. K. Willardson and A. C. Beer (Academic, New York, 1985), Vol. 22, Part D.
[21] F. Capasso, W. T. Tsang, A. L. Hutchinson, and G. F. Williams, Appl. Phys. Lett. **40**, 38 (1982).
[22] F. Capasso, W. T. Tsang and G. F. Williams, IEEE Trans. Electron Devices **ED-30**, 381 (1983).
[23] F. Y. Juang, U. Das, Y. Nashimoto, and P. K. Bhattacharya, Appl. Phys. Lett. **47**, 972 (1985).
[24] H. Shichijo and K. Hess, Phys. Rev. B **23**, 4197 (1981).
[25] Y. C. Chang, D. Z.-Y. Ting, J. Y. Tang, and K. Hess, Appl. Phys. Lett. **42**, 26 (1983).
[26] M. V. Fischetti and S. E. Laux, Phys. Rev. B **38**, 9721 (1988).
[27] K. F. Brennan, D. H. Park, K. Hess, and M. A. Littlejohn, J. Appl. Phys. **63**, 5004 (1988).
[28] K. F. Brennan, N. Mansour, and Y. Wang, Comput. Phys. Commun. **67**, 73 (1991).
[29] B. K. Ridley, *Quantum Processes in Semiconductors*, 2nd ed. (Oxford University Press, Oxford, 1988).
[30] F. H. Pollak, C. W. Higginbotham, and M. Cardona, J. Phys. Soc. Jpn. **21**, 20 (1966).
[31] M. G. Burt, S. Brand, C. Smith, and R. A. Abram, J. Phys. C **17**, 6385 (1984).
[32] C. Jacoboni and P. Lugli, *The Monte Carlo Method for Semiconductor Device Simulation* (Springer, Vienna, 1989).
[33] G. E. Bulman, V. M. Robbins, K. F. Brennan, K. Hess, and G. E. Stillman, IEEE Electron. Device Lett. **EDL-4**, 181 (1983).
[34] K. Brennan and K. Hess, Solid-State Electron. **27**, 347 (1984).
[35] K. Brennan, IEEE J. Quantum Electron. **QE-24**, 2001 (1988).

Theoretical investigation of wave-vector-dependent analytical and numerical formulations of the interband impact-ionization transition rate for electrons in bulk silicon and GaAs

Ján Kolník, Yang Wang, Ismail H. Oğuzman, and Kevin F. Brennan
School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0250

(Received 21 March 1994; accepted for publication 8 June 1994)

The electron interband impact-ionization rate for both silicon and gallium arsenide is calculated using an ensemble Monte Carlo simulation with the expressed purpose of comparing different formulations of the interband ionization transition rate. Specifically, three different treatments of the transition rate are examined: the traditional Keldysh formula, a new **k**-dependent analytical formulation first derived by W. Quade, E. Scholl, and M. Rudan [Solid State Electron. 36, 1493 (1993)], and a more exact, numerical method of Y. Wang and K. F. Brennan [J. Appl. Phys. 75, 313 (1994)]. Although the completely numerical formulation contains no adjustable parameters and as such provides a very reliable result, it is highly computationally intensive. Alternatively, the Keldysh formula, although inherently simple and computationally efficient, fails to include the **k** dependence as well as the details of the energy band structure. The **k**-dependent analytical formulation of Quade and co-workers overcomes the limitations of both of these models but at the expense of some new parameterization. It is found that the **k**-dependent analytical method of Quade and co-workers produces very similar results to those obtained with the completely numerical model for some quantities. Specifically, both models predict that the effective threshold for impact ionization in GaAs and silicon is quite soft, that the majority of ionization events originate from the second conduction band in both materials, and that the transition rate is **k** dependent. Therefore, it is concluded that the **k**-dependent analytical model can qualitatively reproduce results similar to those obtained with the numerical model yet with far greater computational efficiency. Nevertheless, there exist some important drawbacks to the **k**-dependent analytical model of Quade and co-workers: These are that it does not accurately reproduce the quantum yield data for bulk silicon, it requires determination of a new parameter, related physically to the overlap integrals of the Bloch state which can only be adjusted by comparison to experiment, and fails to account for any wave-vector dependence of the overlap integrals. As such the transition rate may be overestimated at those points for which "near vertical," small change in **k**, transitions occur.

I. INTRODUCTION

The theoretical study of interband impact ionization has been greatly aided by the advent of numerical methods. Owing to the complexity of the energy band structure at high energies, at which impact ionization is initiated, as well as the complicated carrier-phonon scattering mechanisms, simplified analytical formulations, such as the early theories of Shockley,[1] Wolff,[2] and Baraff[3] have limited validity. Numerical models for calculating the ionization rate were then advanced based on the Monte Carlo method;[4,5] however, these approaches utilized parabolic or nonparabolic analytical energy bands which are of questionable validity at high carrier energies. The full details of the energy band structure were first accounted for in the Monte Carlo model of Shichijo and Hess.[6] In their model the dynamics of the electrons in bulk GaAs were simulated within the first conduction band, calculated based on an empirical pseudopotential model. Later, Fischetti and Laux[7] developed a more advanced Monte Carlo simulator for studying impact ionization which improved the numerical accuracy as well as incorporated transport within higher conduction bands.

In order to calculate the overall impact ionization rate using the Monte Carlo technique, it is necessary to formulate an expression for the impact ionization transition rate. This formulation is then incorporated into the Monte Carlo simulator and is treated as an additional scattering mechanism. An impact-ionization event is chosen stochastically in the usual way[8] through the use of a random number based on the relative magnitude of the ionization transition rate compared to competing phonon scattering events. Although there has been recent work on formulating the ionization transition rate using a higher-order quantum-mechanical perturbation theory expansion,[9–11] the impact-ionization transition rate is more typically determined from use of Fermi's golden rule. Fermi's golden rule can be expressed as[12]

$$W_{ii}=\int\frac{2\pi}{\hbar}|M|^2\delta(E_f-E_i)\delta(\Delta\mathbf{k})dS_f, \qquad (1)$$

where dS_f represents integration over all the final states, E_i and E_f are the energies of the initial and final states, respectively, $\Delta\mathbf{k}$ is the momentum change during the interaction. The matrix element M in Eq. (1) is formed using both the direct M_d and exchange M_e terms following Ridley[12] as

$$|M|^2=|M_d|^2+|M_e|^2+|M_d-M_e|^2, \qquad (2)$$

where

0021-8979/94/76(6)/3542/10/$6.00 © 1994 American Institute of Physics

$$M_d=\frac{e^2}{\epsilon V}\,\frac{I(\mathbf{k}_1,\mathbf{k}_1')I(\mathbf{k}_2,\mathbf{k}_2')}{|\mathbf{k}_1'-\mathbf{k}_1|^2+\lambda^2}\,,$$
$$M_e=\frac{e^2}{\epsilon V}\,\frac{I(\mathbf{k}_1,\mathbf{k}_2')I(\mathbf{k}_2,\mathbf{k}_1')}{|\mathbf{k}_2'-\mathbf{k}_1|^2+\lambda^2}\,. \tag{3}$$

In Eq. (3), $\mathbf{k}_1,\mathbf{k}_1'$ are the wave vectors of the incident electron before and after the interaction, $\mathbf{k}_1$, $\mathbf{k}_2'$ represent the states of the secondary electron-hole pair after the collision, λ is the static screening factor, and e is the electronic charge; V is the crystal volume, and ϵ is the dielectric function. Quantities $I(\mathbf{k}_1,\mathbf{k}_1'),I(\mathbf{k}_2,\mathbf{k}_2'),I(\mathbf{k}_1,\mathbf{k}_2'),I(\mathbf{k}_2\mathbf{k}_1')$ are overlap integrals.[12] Following Ridley,[12] the term $|M_d-M_e|^2$ is neglected and the direct and the exchange terms are assumed to be equal.

In most previous Monte Carlo simulations (e.g., Refs. 6 and 13) the impact-ionization transition rate is calculated using the Keldysh formula,[14] which can be obtained from Eq. (1) by adopting several simplifying assumptions: These are that the transition is calculated assuming a direct semiconductor with parabolic bands for all of the carriers, the overlap integrals are considered to be constant and the denominator in the matrix element in Eq. (3) is also taken as a constant with thresholds values of $\mathbf{k}_1,\mathbf{k}_1'$. This yields the very well-known quadratic dependence of the transition rate W_{ii} on the energy E of the initiating particle, usually expressed as

$$W_{ii}=pW(E_{th})\left(\frac{E-E_{th}}{E_{th}}\right)^2, \tag{4}$$

where E_{th} is the threshold energy and $W(E_{th})$ is the total phonon scattering rate at threshold. E_{th} and the prefactor p cannot generally be determined from first principles; instead, they are determined through comparison of the calculations to experimental data. Aside from this obvious limitation of the Keldysh formula, the neglect of an accurate accounting of the high-energy region of the band structure also results in the failure to properly assess the **k** dependence of the ionization transition rate itself. Recent theoretical studies have indicated that there is a significant deviation in the ionization transition rate with respect to the initiating carrier's **k** vector in different materials.[15,16]

Several alternative approaches to the Keldysh formula for calculating the interband impact ionization rate have been recently presented.[15–21] In each of these methods, the interband impact-ionization transition rate is evaluated numerically and can be then incorporated into an ensemble Monte Carlo calculation. Although these approaches presently offer a far more accurate treatment of impact ionization than the Keldysh formula, they require extensive numerical computation. In the approach adopted by Wang and Brennan[15,16] the transition rate is determined directly from Fermi's golden rule by numerically integrating Eq. (1) at different **k** points within the first Brillouin zone. Their analysis includes the direct calculation of the overlap integrals from the numerically generated wave functions using a $\mathbf{k}\cdot\mathbf{p}$ method. The transition rate is evaluated by integrating over several million final states directly evaluating the overlap integrals in each case. This is particularly computationally intensive since the overlap integrals must be evaluated for each final state. Sano and Yoshii[17] have used a less computationally intensive scheme. In their method, the three-dimensional **k** space incorporating the reduced zone is partitioned into small cubes. The matrix elements within each cube are assumed to be equal alleviating the need to evaluate the overlap integrals for each possible state. In this way, the number of computations can be greatly reduced at the expense of increased memory requirements. However, we have found that the overlap integrals vary in a nonpredictive manner from point to point in our $\mathbf{k}\cdot\mathbf{p}$ calculations and cannot generally be assumed to be the same even for states very close in **k** space. For this reason and memory storage limitations, in the model of Wang and Brennan[15,16] the overlap integrals are evaluated for every final state. Nevertheless, all of the numerical approaches indicate that the transition rate is dependent on the initiating electron wave vector.

Recently, another analytical treatment of expression (1), which also includes the **k**-vector dependence of the impact-ionization transition rate, was presented by Quade, Scholl, and Rudan.[22] They calculated the impact-ionization transition rate for the general case of an arbitrarily shaped energy band for the impact-ionizing conduction electron, and three anisotropic parabolic bands for the final states with their extrema located at different points in the **k**-vector space. The assumption of the parabolic energy bands for the final states is acceptable since the final states are generally at low energy near the band minimum where the parabolic approximation is typically good. Other approximations used in the derivation include nondegeneracy and a constant value for the overlap integrals. Although the formulation developed by Quade and co-workers[22] contains some parameterization due to the fact that the overlap integrals are not evaluated but are assumed to be constant parameters, it offers a far more computationally efficient means of evaluating the ionization transition rate with the inclusion of the **k** dependence than the numerical method of Wang and Brennan.[15,16] Subsequently, it is of interest to examine how the **k**-dependent, analytical model of Quade and co-workers[22] compares to the numerical model of Wang and Brennan.[15,16]

In this article, we incorporate the formulation of the impact-ionization transition rate of Quade and co-workers[22] within an ensemble Monte Carlo simulator to determine the electron impact-ionization rate in bulk silicon and GaAs. The impact-ionization rate calculated in this way is then compared to experimental data as well as calculations made in an otherwise identical Monte Carlo simulator using the Keldysh formula and the direct, **k**-dependent transition rate. The nature of the ionization rate is further probed using the Quade and co-workers' model to determine if it yields similar physical results to that of the direct, numerical model. Specifically, the transition rate within each band is calculated and the percentage of ionization events originating from each conduction band is determined. The number density distributions and the quantum yield as a function of energy are determined and compared between the different theoretical models as well. In Sec. II the details of the theoretical models are reviewed. The calculated results are presented in Sec. III and conclusions are drawn in Sec. IV.

II. MODEL DESCRIPTION

The total impact-ionization rate as a function of inverse electric field is calculated using an ensemble Monte Carlo simulation, based on the original work of Shichijo and Hess.[6] The details of the simulator have been extensively described elsewhere;[15,16] therefore, only the main features are outlined here. The simulation includes the full details of the first and second conduction bands for silicon or gallium arsenide. The scattering rate in either case is determined using the improved phonon scattering technique of Chang *et al.*[23] and the improved band-structure interpolation technique of Fischetti and Laux.[7] The total phonon scattering rate is determined in the following way. Within the low-energy region, the scattering rate is calculated from Fermi's golden rule for all the relevant mechanisms present.[24] Within the high-energy range, deformation potential scattering is assumed to be the dominant scattering mechanism. The scattering rate in this region is obtained by integrating over the final density of states, calculated from the numerically generated band structures, including collision broadening of the final state. The deformation potential is assumed constant and is selected to match the scattering rate calculated from Fermi's golden rule at a low specific energy. Interband electronic transitions are enabled via the action of the deformation potential scattering as described in Ref. 15. The actual scattering rates used in these calculations have been published previously in Ref. 15 for silicon and Ref. 16 for GaAs.

The impact-ionization transition rate is formulated in three different ways and is then incorporated into the Monte Carlo simulator described above. The different transition rate formulations used are a direct, numerical method previously described in detail in Refs. 15 and 16, the Keldysh formula described in detail in Refs. 6 and 7, and a new **k**-dependent analytical formulation derived by Quade and co-workers[22] which is summarized below.

The **k**-dependent analytical formulation of the impact-ionization transition rate, hereafter referred to as the Quade formula, predicts the ionization transition rate for an initiating electron of wave vector $\mathbf{k}_1$, colliding with another electron in the valence band, to be[22]

$$W_{\mathrm{ii}}(\mathbf{k}_1)_{\alpha,\beta,\gamma}=\int_{-\infty}^{K_1^2} d\xi\, W_{\mathrm{ii}}(K_1,\xi)_{\alpha,\beta,\gamma}\,\delta[g(\mathbf{k}_1)-\xi], \tag{5}$$

where

$$W_{\mathrm{ii}}(K_1,\xi)_{\alpha,\beta,\gamma}=\frac{1}{\tau}\left[\frac{1}{2}\left(\frac{K_1}{\sqrt{\xi_\lambda}}+\frac{\sqrt{\xi_\lambda}}{K_1}\right)-1\right]\Theta(K_1^2-\xi), \tag{6}$$

$$\frac{1}{\tau}=\frac{1}{4\pi^2}\,\frac{\sqrt{\mu}}{\Pi_{i=1}^{3}(\alpha_i+\beta_i)^{1/2}}\left(\frac{e^2}{\epsilon}\right)^2\frac{m_\delta F}{\hbar^3}, \tag{7}$$

$$\mu_i=\frac{\alpha_i\beta_i+\alpha_i\gamma_i+\beta_i\gamma_i}{\alpha_i+\beta_i}, \tag{8}$$

$$g(\mathbf{k}_1)=\frac{1}{\mu}\left(E_0+\sum_{i=1}^{3}\frac{\alpha_i\beta_i}{\alpha_i+\beta_i}(Z_{\alpha,i}-Z_{\beta,i})^2\right.$$
$$\left.+\sum_{i=1}^{3}\gamma_i(k_{1,i}^L-Z_{\gamma,i})^2-E(\mathbf{k}_1^L)\right), \tag{9}$$

$$K_{1,i}=\frac{1}{\sqrt{\mu\mu_i}}\left(\gamma_i(k_{1,i}^L-Z_{\gamma,i})-\frac{\alpha_i\beta_i}{\alpha_i+\beta_i}(Z_{\alpha,i}-Z_{\beta,i})\right), \tag{10}$$

and

$$\xi_\lambda=\tfrac{1}{2}\left[\sqrt{(\lambda^2-\xi)^2+4K_1^2\lambda^2}-(\lambda^2-\xi)\right]. \tag{11}$$

In the above set of equations, α, β, and γ represent a parabolic valence band and two possibly different parabolic conduction bands for the final states of the hole and electrons after the impact ionization event, respectively. The anisotropic bands are parametrized by ratios of effective masses $\alpha_i=m_\delta/m_{\alpha,i}$, $\beta_i=m_\delta/m_{\beta,i}$, $\gamma_i=m_\delta/m_{\gamma,i}$, where i labels the Cartesian coordinate, and the extremal points of each band are positioned at $\mathbf{Z}_\alpha$, $\mathbf{Z}_\beta$, $\mathbf{Z}_\gamma$; m_δ is an isotropic effective mass of the conduction band for the initiating electron. F represents the product of the squared overlap integrals and is taken as constant for each initiating band but may vary between bands. μ stands for the geometrical average of the quantities μ_1, μ_2, μ_3 defined by Eq. (8). The arbitrarily shaped function $E(\mathbf{k}_1)$ represents the energy of the initiating electron, E_0 represents the gap plus offsets of the extremal points of the bands α, β, γ. $\mathbf{k}_1^L=\mathbf{k}_1-\mathbf{L}$, where $\mathbf{L}$ is a reciprocal lattice vector and Θ is the Heaviside function. Equation (5) is derived assuming a static dielectric constant. Although recent work[17] has shown that a full wave-vector, frequency-dependent dielectric function can modify the ionization transition rate, its incorporation in the Quade formula would not lead to an analytical expression. For this reason, a constant dielectric function is used within the Quade formulation. To facilitate comparison of the Quade model to other models, a constant dielectric function is used within the numerical and Keldysh formula models as well.

The total impact-ionization scattering rate for the initiating particle with wave vector $\mathbf{k}_1$ is calculated taking into account all of the possible final states within the full Brillouin zone. The total impact-ionization transition rate can then be expressed as

$$W_{\mathrm{ii}}(\mathbf{k}_1)=\sum_{\alpha}\sum_{\beta}\sum_{\gamma}W_{\mathrm{ii}}(\mathbf{k}_1)_{\alpha,\beta,\gamma}, \tag{12}$$

where the summation goes over all the valence bands α and all possible valleys β, γ in the first and second conduction bands. The summations in Eq. (12) are taken so as to include both direct and exchange terms.

The material parameters used in the calculations have been taken from Refs. 25–28 for silicon and Refs. 29–32 for gallium arsenide. For the case of the second conduction band in both materials, the longitudinal m_l and transverse m_t effective masses for the X valleys have been obtained from the numerically calculated band structures determined by a $\mathbf{k}\cdot\mathbf{p}$ calculation. The longitudinal and transverse masses used for silicon and gallium arsenide in the second conduction band

TABLE I. Energies of symmetry points (relative to the top of the valence band, in eV) and values of relative effective masses used in the calculations of the impact-ionization transition rate in silicon and gallium arsenide.

	Silicon	Gallium arsenide
$E(\Gamma_{CI})$	3.42[a]	1.423[b]
$E(X_{CI})$	1.12[a]	1.898[b]
$E(L_{CI})$	1.92[a]	1.706[b]
$E(X_{CII})$	1.14[c]	1.973[c]
$E(\Gamma_{SO})$	−0.044[d]	−0.340[b]
$m_t(X_{CI})$	0.19[e]	0.27[f]
$m_l(X_{CI})$	0.92[e]	1.98[f]
$m_t(X_{CII})$	0.21[c]	0.23[c]
$m_l(X_{CII})$	0.29[c]	0.40[c]
$m(L_{CI})$	0.284[e]	0.18[f]
$m(\Gamma_{CI})$	...	0.067[f]
$m(\Gamma_{HH})$	0.537[d]	0.45[f]
$m(\Gamma_{LH})$	0.153[d]	0.082[f]
$m(\Gamma_{SO})$	0.234[d]	0.170[f]

[a] Reference 32.
[b] Reference 30.
[c] Values obtained from the numerically calculated band structure.
[d] Reference 26.
[e] Reference 25.
[f] Reference 31.

are reported in Table I. In the case of L valleys, Γ valleys, and all the valence bands, the isotropic effective masses have been used and their values are also listed in Table I, as well as the energies of symmetry points relative to the top of the valence band.

The interband impact-ionization transition rate calculated using the approach based on Quade and co-workers' results assuming that the overlap integrals are simply unity is presented in Figs. 1(a) and 1(b) for the first and second conduction bands in silicon, and in Figs. 2(a) and 2(b) for the first and second conduction bands in gallium arsenide, respectively. Comparison of Figs. 1 and 2 to similar figures obtained by Wang and Brennan[15,16] using the completely numerical technique in which the overlap integrals are evaluated directly shows good qualitative agreement between the two approaches. For bulk silicon the primary difference, aside from the overall magnitude, is that the number of points for which a nonzero transition rate has been calculated is higher using the completely numerical approach. A similar effect can be seen in the case of the first and second conduction bands in gallium arsenide. Again notice that there is a significant difference in the number of nonzero transition rate points between the two models (data calculated using the numerical approach can be found in Ref. 16) for the first conduction band; however, the rate from these points is relatively small compared to the competing scattering rates.

Further examination of Figs. 1 and 2 reveals that the transition is greater for electrons originating from the second conduction band than from the first conduction band, which corresponds to the results based on the numerical calculations.[15,16] This is particularly true for GaAs. As is discussed below, due to the fact that the transition rate is greatest for the second conduction band, most ionization events occur from electrons which originate from within the second conduction band.

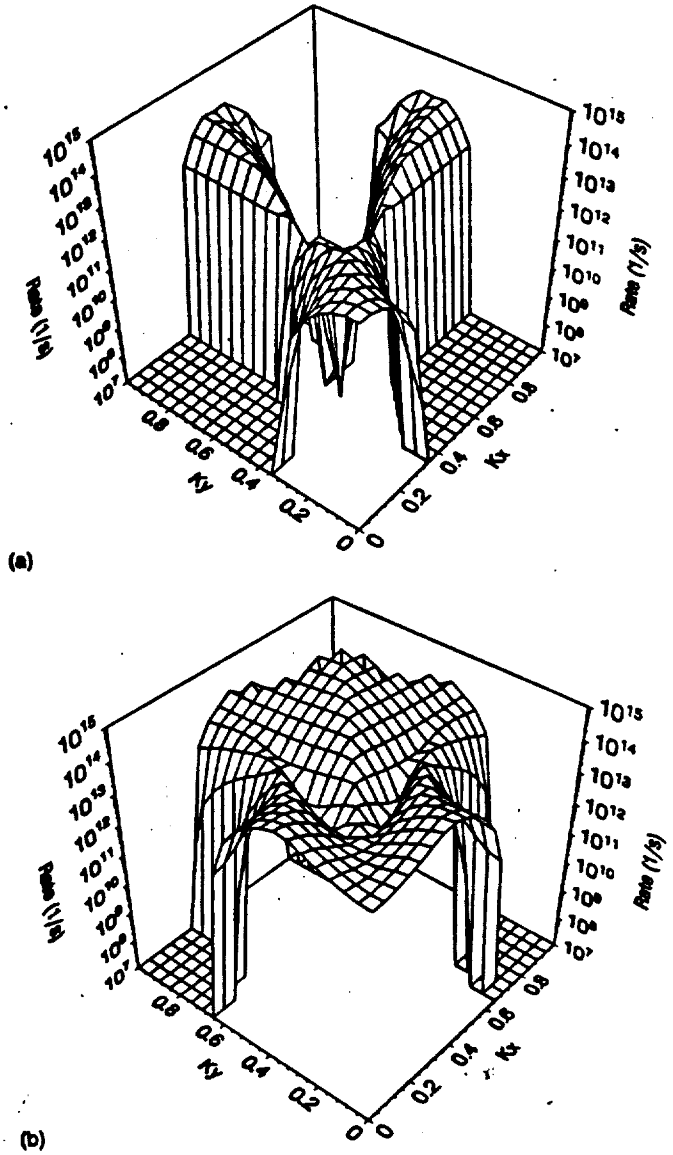

FIG. 1. Calculated interband impact-ionization transition rate for the first and second conduction band in bulk silicon as a function of k_x and k_y (k_z is fixed at 0.0) within the reduced zone of the first Brillouin zone obtained by the analytical formulation based on the Quade model. Initiating electron is in (a) the first conduction band and (b) the second conduction band.

III. RESULTS

The impact-ionization rate as a function of inverse electric field, calculated using the ensemble Monte Carlo simulation, is presented in Figs. 3 and 4 for silicon and gallium arsenide, respectively. In each case, the rate is calculated using the Keldysh formula, the k-dependent numerical formulation, and the Quade formula for the impact-ionization transition rate. The experimentally obtained data are also shown in both figures. The experimental results of van Overstraeten and DeMan,[33] Grant,[34] and Woods, Johnson, and Lambert[35] for electron-initiated impact ionization in silicon are plotted in Fig. 3. These data comprise a representative set of data which cover the full range of experimental measurements for electron-impact ionization in silicon. The experimental measurements of Bulman *et al.*[36] are plotted for gallium arsenide in Fig. 4.

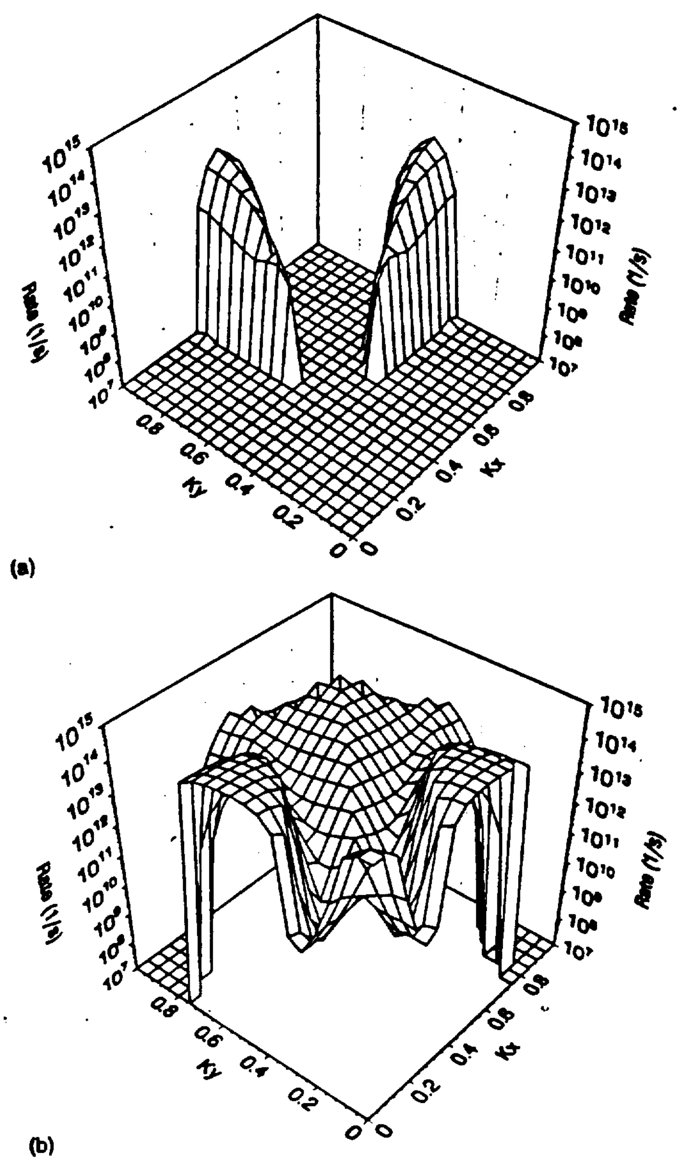

FIG. 2. Calculated interband impact-ionization transition rate for the first and second conduction band in bulk GaAs as a function of k_x and k_y (k_z is fixed at 0.0) within the reduced zone of the first Brillouin zone obtained by the analytical formulation based on the Quade model. Initiating electron is in (a) the first conduction band and (b) the second conduction band.

In the case of silicon, good agreement with the data of Woods and co-workers is obtained for the k-dependent numerical formulation, as was also reported previously.[15] No adjustable parameters for the impact-ionization transition rate are used in this case. Very similar results can be obtained using the Quade formulation by choosing the values for the parameter F in Eq. (7) equal to 0.24 for both the first and second conduction bands. F corresponds to an "average" value of the squared overlap integrals. The squared overlap integrals always have value between 0 and 1. In the analytical model of Quade and co-workers[22] and that presented here, F is treated as a parameter which is chosen so as to yield good agreement with experiment. Best agreement with the low range of silicon experimental data occurs for $F=0.24$ given the phonon scattering rate adopted here. The calculated silicon impact-ionization rate obtained using the

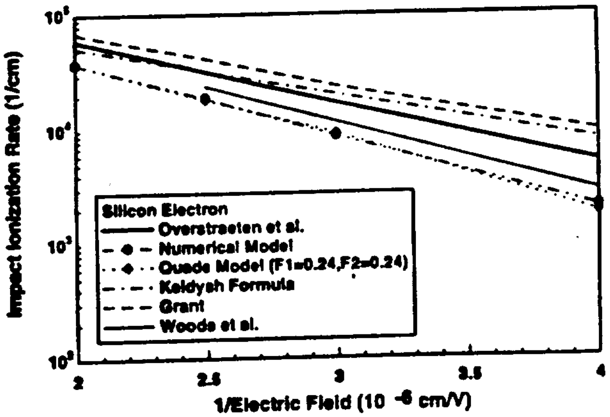

FIG. 3. Calculated and experimental results for the electron impact-ionization coefficients in bulk silicon plotted as a function of inverse electric field. Results of the calculations of the ionization rate including the Keldysh formula, k-dependent numerical, and Quade model are presented. Experimental results of Refs. 33, 34, and 35 are also shown. F1 and F2 represent the values of the averaged squared overlap integrals used in the Quade model for bands 1 and 2, respectively.

Keldysh formula with $p=0.02$ and $E_{th}=1.13$ eV is also shown in Fig. 3. A value of p of about 0.02 is typically considered to correspond to a soft threshold. In contrast to both k-dependent models, better agreement with the higher range of experimental measurements can be seen for this particular choice of threshold and p; however, a different choice of the parameter F within the Quade formulation would also provide a better fit to the higher range of experimental data as well.

In gallium arsenide the dependence of the impact-ionization rate on the inverse electric field obtained using the numerical, the Quade models with $F=0.05$ and 0.005 for the first and second conduction band, respectively, and the Keldysh formula with $p=0.02$, $E_{th}=1.85$ eV for the impact-ionization transition rate, shows in all three cases fairly good agreement with the experimentally measured data

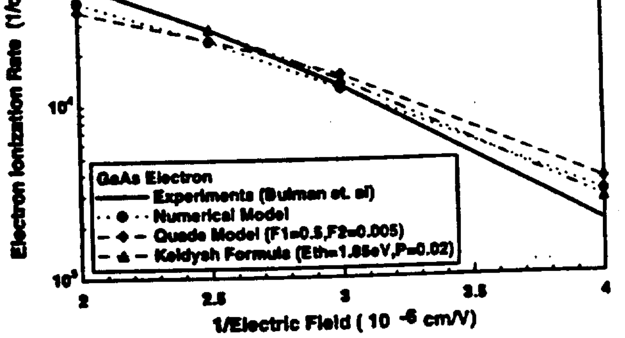

FIG. 4. Calculated and experimental results for the electron impact-ionization coefficients in bulk gallium arsenide plotted as a function of inverse electric field, using the Keldysh formula, k-dependent numerical, and Quade model. Experimental results shown are those of Bulman *et al.* (Ref. 36). F1 and F2 represent the values of the averaged squared overlap integrals used in the Quade model for bands 1 and 2, respectively.

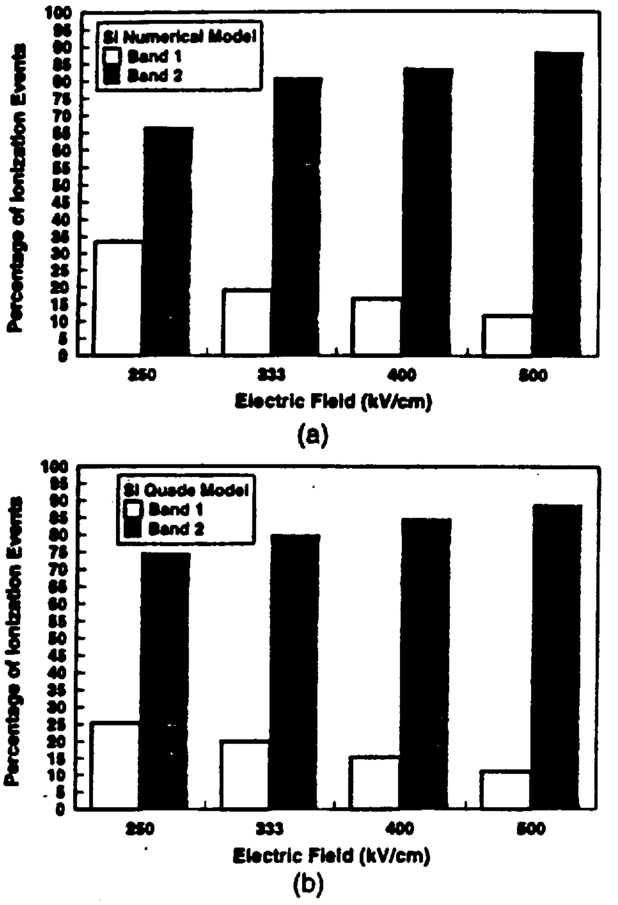

FIG. 5. Histogram showing the percentage of ionization events originating from the first and second conduction bands in silicon for (a) the k-dependent numerical formulation and (b) the Quade formulation.

(Fig. 4). It should be noted that in the case of the approach based on the Quade formula, the selection of the fitting factor F for the first conduction band essentially does not play any role due to the negligible impact-ionization transition rate of the electrons in the first conduction band with respect to the second conduction band.

The percentage of ionization events originating from the first and the second conduction band in bulk silicon as a function of applied electric field is plotted in Fig. 5(a) for the numerical model and in Fig. 5(b) for the model based on the Quade formula. At the applied electric fields investigated here the impact-ionization events originate predominantly in the second band in both cases. Therefore, on the basis of these reports, it appears that most electron-initiated ionization events originate from electrons within the second conduction band. The number of events originating in the first conduction band is found to be higher for the numerical model than for the model using the Quade formula at the field of 250 kV/cm, while for higher fields this difference vanishes. This slightly different behavior of the two models at low fields is caused by the fact that the threshold energy for impact ionization is about 1.32 eV in the Quade model for the particle originating in the first conduction band and 2.11 eV for the second conduction band, while for the numerical model the threshold energy value is not well defined.

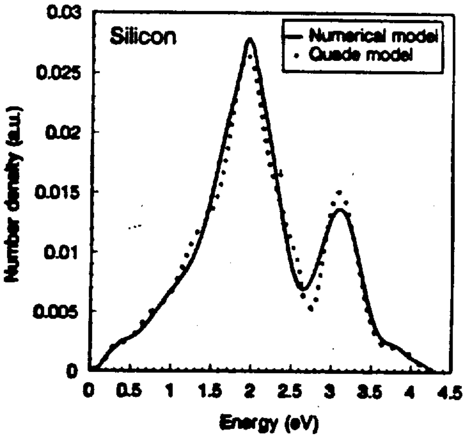

FIG. 6. The electron number density function in silicon, defined as the product of the density-of-states function and the electron distribution function, at an applied electric field of 500 kV/cm, calculated using the k-dependent numerical model and the Quade model.

Instead, in the numerical model ionization is possible at all states with energy greater than the band gap which are found to satisfy the momentum and energy conservation conditions. Therefore, there is no sharp threshold energy, other than the band gap, below which ionization cannot occur and above which ionization does occur. Thus, the contribution of the first conduction band at low fields can be expected to be higher when the numerical model is used.

The effect of the second conduction band on the impact-ionization rate in gallium arsenide is even more important than in silicon. The impact-ionization transition rate for electrons in the first conduction band is much lower than the rate within the second conduction band in both models as can be seen from Figs. 2(a) and 2(b) and from Ref. 16. As a result, virtually all the electrons which impact ionize originate within the second conduction band. The effect is even stronger using the Quade formulation than the numerical formulation. Though the value of F used in the analytical formulation is unknown *a priori*, i.e., without first comparing the calculations to experiment, it should be noted that F can never be greater than 1. Therefore, given that the transition rate of electrons within the first conduction band is several orders of magnitude lower than from the second conduction band with $F=1$, as shown by Figs. 2(a) and 2(b), it is clear that the first conduction band plays no significant role in impact ionization in bulk GaAs for reasonable values of F. Hence, in either the numerical or analytical k-dependent models the predominant source of ionizing electrons in bulk GaAs is carriers within the second conduction band.

The electron number density function defined as the density-of-states function multiplied by the electron distribution function, is shown in Fig. 6 for silicon at an applied electric field of 500 kV/cm for both the numerical k-dependent model and the Quade model. In both cases, the number of electrons which survive to high energies is relatively high. The second peak in each of the figures corresponds to the electron population in the second conduction band and is slightly larger for the Quade model. Both models essentially confirm previous results[37–39] that the threshold in

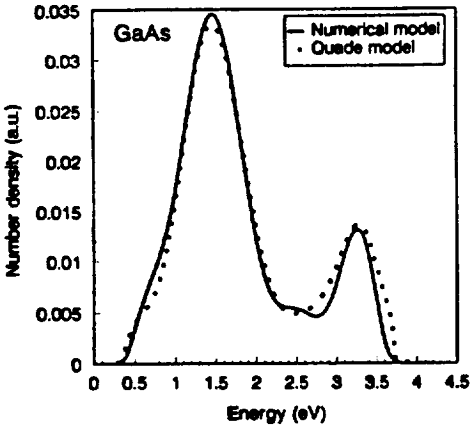

FIG. 7. The electron number density function in gallium arsenide, defined as the product of the density-of-states function and the electron distribution function, at an applied electric field of 500 kV/cm, calculated using the k-dependent numerical model and the Quade model.

silicon is exceedingly soft and that the electrons can reach energies substantially higher than the band gap before they impact ionize.

The electron number density functions in gallium arsenide are plotted in Fig. 7 calculated using the k-dependent numerical and Quade models for the applied electric field of 500 kV/cm. In both cases the distribution functions are similar and the effect of the second conduction band (the second peak at higher energies) is again slightly larger in the Quade model. Notice that the number distribution remains substantially large at high carrier energies indicating an exceedingly soft effective threshold for impact ionization in GaAs. Again these results are consistent with previously reported observations.[29]

The energy-dependent impact-ionization transition rate, derived from the k-dependent rate in a manner described in Ref. 15, is plotted in Fig. 8 for the numerical k-dependent formulation and the Quade formula, respectively, for silicon.

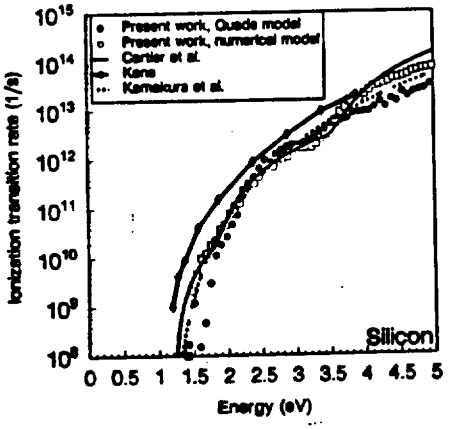

FIG. 8. Calculated impact-ionization transition rate in silicon as a function of electron energy measured from the conduction-band minimum. The energy-dependent rate is calculated by performing a further integration of the k-dependent rate following the approach explained in Ref. 15 for the numerical and Quade models; the results of Kane (Ref. 40) and Cartier *et al.* (Ref. 41) and Kamakura *et al.* (Ref. 19) are also shown.

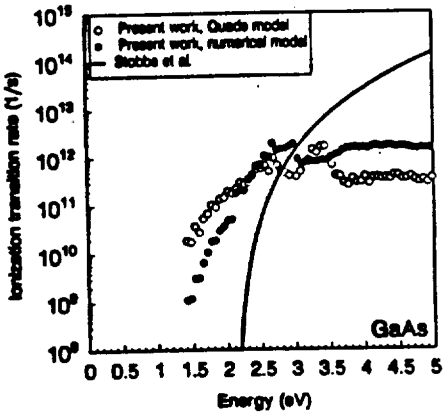

FIG. 9. Calculated impact-ionization transition rate in gallium arsenide as a function of electron energy measured from the conduction-band minimum using the k-dependent numerical model and the Quade model. The energy-dependent rate is calculated by performing a further integration of the k-dependent rate following the approach explained in Ref. 15. The energy-dependent rate obtained by Stobbe and co-workers (Ref. 21) is also included.

The results of Kane,[40] Kamakura *et al.*,[19] and the recently published ionization transition rate of Cartier *et al.*[41] determined from a comparison of Monte Carlo simulations and soft-x-ray photoemission spectroscopy (XPS) measurements are shown as well. The numerically calculated transition rate is almost identical with that of Cartier *et al.* up to an energy of 4 eV, and seems to be closer to the results of Kane in the high-energy region. Recently, also, Kunikiyo *et al.*[18] have performed similar calculations of the silicon interband impact-ionization transition rate. Their calculations are in excellent agreement as well with the transition rate of Cartier *et al.*[41] The transition rate obtained on the basis of the Quade formula agrees well with the numerical model between 2 and 4 eV, while a lower rate is predicted in the low- and high-energy regions.

The energy-dependent impact-ionization transition rate for gallium arsenide, calculated using the numerical approach of Wang and Brennan[16] and the Quade formula, respectively, are shown in Fig. 9. Stobbe and co-workers[21] have also recently calculated the impact-ionization transition rate in bulk GaAs. Their calculations, based on the fitting formula quoted in Ref. 21, are included in Fig. 9 for comparison. Inspection of Fig. 9 shows that the maxima of the transition rate, based on both the numerical model of Wang and Brennan[16] and the Quade model, occurs near 3 eV in GaAs. The maximum region is more pronounced for the Quade formulation than in the case of Wang and Brennan's numerical model. The origin of the maximum can be possibly understood within the analytical formulation, since the transition rate involving each valley can be individually evaluated. The maximum in the transition rate is due to the collision of an electron within the second conduction band with an electron in the heavy-hole band, after which the electrons occupy states either both in the Γ valley (the lower peak below 3.0 eV) or one is in the Γ valley and the other in the *L* valley (the higher peak close to 3.4 eV). In both cases, a transition of the initiating electron from the second conduction band to the first one with only a slight change of the **k** vector (a near-vertical transition, as explained in Ref. 42) contributes to the final values of the transition rate. Near-vertical transitions have a significantly greater rate owing to the fact that the denominator in Eq. (3) becomes very small under these conditions.[42] This region may be somewhat unrealistically pronounced in the Quade formulation due to the fact that the value of the overlap integrals is taken as an averaged constant. The actual values of the overlap integrals in this region may be quite small thereby reducing the rate if calculated. It is much more difficult to isolate the physical origin of the maximum in the numerical model. In this case, a similar classification would be questionable, mainly due to the uncertainty in how the "valleys" are defined in the real band structure. Inspection of Fig. 9 also shows that the study of Stobbe and co-workers[21] predicts a higher ionization transition rate from the results presented here. The cause of this discrepancy may be due to the inclusion of a q-dependent dielectric function in Stobbe and co-workers' formulation as well as differences in the energy band structure utilized in the two calculations. In Stobbe and co-workers' work the band structure is calculated using an empirical pseudopotential method in contrast to the $\mathbf{k}\cdot\mathbf{p}$ method adopted in the present work.

To further investigate the impact-ionization transition rate formulations, the energy dependence of the quantum yield was also calculated for both the numerical and Quade models. The quantum yield is defined as the average number of impact-ionization events caused by a high-energy injected electron until its kinetic energy relaxes below the ionization threshold through scattering and/or ionization events. While the impact-ionization coefficient strongly depends on the energy distribution function and therefore mainly on the inelastic electron-phonon scattering rate, the quantum yield is almost a linear function of the ratio of the ionization transition rate to the electron-phonon scattering rate.[41] The quantum yield provides a useful means of assessing the relative hardness or softness of the impact-ionization threshold and enables an additional, independent evaluation of the appropriateness of the model used.

To the authors' knowledge, the only presently existing experimentally based data for the quantum yield apply to bulk silicon. Calculations of the quantum yield in silicon are presented in Fig. 10 using both the numerical and Quade models. The data of Cartier *et al.*,[41] derived from a comparison of Monte Carlo simulations and XPS measurements, which were shown to be in good agreement with other measurements,[41] are also plotted in Fig. 10 for comparison. While fairly good agreement over the full energy range between the model based on the numerically calculated ionization rate and the data from Ref. 41 can be seen, the Quade formula predicts a quantum yield which is substantially lower in the low-energy region. Nevertheless, the Quade formula results correspond fairly well to the other data above 2.5 eV. These results can be understood on the basis of the threshold energy for impact ionization. In the case of the Quade formula the threshold energy is well defined and is substantially higher than the value of the gap. Subsequently the ionization-transition rate in this region is lower for the Quade model, as can be seen in Fig. 8, which results in a lower quantum yield. Conversely, in the numerical model, the effective threshold is simply the energy gap and subsequently ionization events can possibly occur at low energies in this model leading to a higher quantum yield.

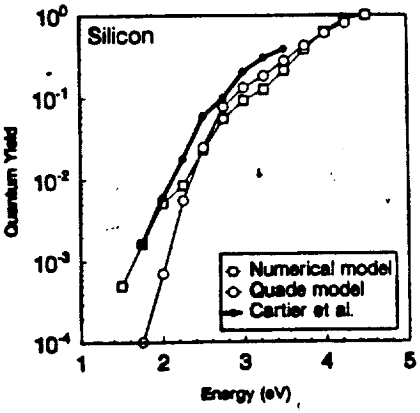

FIG. 10. Calculated quantum yield as a function of electron energy in silicon measured from the conduction-band minimum, obtained using the **k**-dependent numerical model and the Quade model. Results of Cartier *et al.* (Ref. 41) are also shown.

Although no experimental data for bulk GaAs are presently available, it is nevertheless instructive to compare the calculated quantum yield dependencies on energy using the numerical and Quade models. The calculated quantum yield versus electron energy in GaAs using the numerical and Quade models is shown in Fig. 11. Comparison of the calculations reveals good correspondence between these two models in the energy range from 2.5 to 3.5 eV, while at higher energies the Quade model [with the particular choice of the prefactor F appearing in Eq. (7), as described above] predicts a lower quantum yield than the numerical model. At lower energies the Quade model predicts a higher quantum yield than the numerical model. This discrepancy can be understood as follows. An identical phonon scattering rate is used in both models of the ionization rate. As stated by Cartier *et al.*[41] the quantum yield depends linearly upon the ratio of the impact-ionization transition rate to the phonon scattering

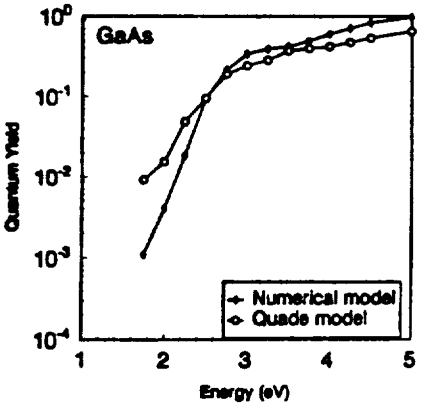

FIG. 11. Calculated quantum yield as a function of electron energy in gallium arsenide measured from the conduction-band minimum, obtained using the **k**-dependent numerical model and the Quade model.

rate. Subsequently, the quantum yield within each model, given that the scattering rates are identical, depends then simply upon the ratio of the transition rates and will follow the same trends as shown in Fig. 9.

The value of the overlap integrals used in the evaluation of the transition rate given by Eqs. (2) and (3) is evaluated directly in the numerical approach. As mentioned above, this requires extensive numerical computation. By approximating these overlap integrals as a constant, the factor F in Eq. (7), a new single parameter is introduced greatly reducing the numerical complexity required to evaluate the transition rate but at the expense of a new unknown parameter. In the calculations presented above, F is adjusted such that the analytical model yields satisfactory agreement with the experimental data. To check to see if the determined values of F are reasonable, it is useful to compare these values to those determined using the sum rule quoted by Ridley.[12] Using the sum rule, the squared overlap integral between the first conduction band and the heavy-hole band in gallium arsenide is determined to be 0.38. More elaborate calculations made by Burt *et al.*[43,44] using wave functions obtained from a $\mathbf{k}\cdot\mathbf{p}$ calculation show that this value is $\sim 10^{-2}$, which is in fair agreement with the results presented here for GaAs, $F=0.005$ for the second conduction band. In the case of silicon, the value $F=0.24$ obtained for the impact ionization originating in both the first and second conduction bands, is higher than the value of $F\sim 0.05$,[45] or $F=0.01$, quoted in Refs. 46 and 47, obtained by theoretical evaluations. It should be noted, though, that a discrepancy between the experimentally determined recombination coefficients for Auger recombination, and the theoretically obtained ones, which also contain the same overlap integrals, was observed in bulk silicon. This discrepancy is thought to be possibly due to an underestimation of the overlap integrals by roughly an order of magnitude.[48]

IV. CONCLUSIONS

In this article the electron interband impact-ionization rate calculated by an ensemble Monte Carlo simulation has been presented for both bulk silicon and gallium arsenide with the expressed purpose of comparing different formulations of the transition rate. Specifically, three different treatments of the transition rate are examined: the traditional Keldysh formula, a new **k**-dependent analytical formulation first derived by Quade and co-workers,[22] and a more exact, numerical method of Wang and Brennan.[15,16] The numerical method of Wang and Brennan[15,16] contains no readily adjustable parameters and is similar to other numerical techniques that are currently attaining prominence. The numerical technique thus serves as a standard to which the Quade result has been compared. Unfortunately, the completely numerical technique, although fundamentally based, requires extensive numerical computation making it somewhat unattractive. As is well known, the Keldysh formula, although commonly used in Monte Carlo calculations of the impact-ionization rate owing to its relative simplicity and computational efficiency, suffers from several limitations as well. Chief among these is the fact that it is derived assuming parabolic energy bands for all of the participating carriers, and includes no **k** dependence in the rate. Additionally, due to its reliance on two adjustable parameters, more than one set of parameters can be used within the Keldysh formula to produce satisfactory agreement with experiment. As a result, the Keldysh formula cannot by itself predict whether a material has a soft or hard ionization threshold. Therefore, there is a need to develop a more computationally efficient transition rate formulation than the direct, numerical technique, yet retain the **k** dependence and details of the band structure in its formulation. In this way, the physical nature of the ionization process can be preserved (soft or hard threshold, etc.) while maintaining computational efficiency. The Quade formula, although it contains one adjustable parameter corresponding physically to the squared overlap integrals of the periodic parts of the Bloch functions, includes the full details of the band structure for the initiating carrier. In this article, we have provided the first comparison of this formula to the numerical technique to determine if this approach yields qualitatively similar results while providing far better computational efficiency.

The calculations presented here show that the model based on the Quade formulation produces similar results to the completely numerical calculations for some quantities, primarily the ionization rate as a function of field. In contrast, to the Keldysh formula, where no **k**-vector dependence of the impact-ionization transition rate is included, the transition rate in the **k**-dependent numerical and analytical (Quade formula) cases is shown to be strongly dependent on the initiating electron **k** vector. It is also shown that both **k**-dependent models predict that the second conduction band dominates the ionization process in GaAs and silicon; however, the quantum yield predicted using the analytical model of Quade and co-workers[22] does not agree closely with the available experimentally derived data of Cartier *et al.*[41] or that predicted using the numerical model of Wang and Brennan.[15,16] As such the Quade formula, although far more complete than the Keldysh formula, may not be a completely reliable replacement for the numerical models in many situations.

Further comparison of the Quade and numerical models reveals that they predict that the effective threshold for impact ionization is relatively soft, implying that the carriers drift to substantially high energies before suffering an ionization event. It is shown that the average energy of the electrons is high when compared to that expected for a hard threshold model. Thus, many of the important qualitative results obtained by adopting the computationally intensive numerical approach, i.e., the dominance of the second conduction band, the nature of the ionization threshold, and the importance of the **k** dependence, can be recovered if the analytical **k**-dependent formulation of Quade and co-workers[22] is used. Use of the Quade formula may not always be fully justifiable, however, since it contains some *ad hoc* parametrization which cannot be ascertained from first principles and it apparently does not reproduce the quantum yield data. Additionally, it is possible that the Quade formula may overestimate the impact-ionization transition rate at those points in the **k** space where near-vertical transitions are allowed due to the failure to include the actual values of the overlap integrals for these transitions.

ACKNOWLEDGMENTS

The authors would like to thank Dr. C. J. Summers and Dr. B. K. Wagner for many helpful discussions. This work was supported in part by the National Science Foundation through a Presidential Young Investigator Award made to K. Brennan, through Bell Northern Research Grant No. I25A3J and by ARPA through contract to NASA, NAGW-2753. Computing support was provided by the Digital Equipment Corp. through Contract No. E21-H36.

[1] W. Schockley, Solid-State Electron **2**, 35 (1961).
[2] P. A. Wolff, Phys. Rev. **95**, 1415 (1954).
[3] G. Baraff, Phys. Rev. **128**, 2507 (1962).
[4] R. C. Curby and D. K. Ferry, Phys. Status Solidi A **15**, 319 (1973).
[5] P. A. Lebwohl and P. J. Price, Solid State Commun. **9**, 1221 (1971).
[6] H. Shichijo and K. Hess, Phys. Rev. B **23**, 4197 (1981).
[7] M. V. Fischetti and S. E. Laux, Phys. Rev. B **38**, 9721 (1988).
[8] C. Jacoboni and P. Lugli, *The Monte Carlo Method for Semiconductor Device Simulation* (Springer, Vienna, 1989).
[9] J. Bude, K. Hess, and G. J. Iafrate, Phys. Rev. B **45**, 10 958 (1992).
[10] W. Quade, F. Rossi, and C. Jacoboni, Semicond. Sci. Technol. **7**, B502 (1992).
[11] P. D. Yoder, J. M. Higman, J. Bude, and K. Hess, Semicond. Sci. Technol. **7**, B357 (1992).
[12] B. K. Ridley, *Quantum Processes in Semiconductors*, 2nd. ed. (Oxford University Press, Oxford, 1988).
[13] J. Y. Tang and K. Hess, J. Appl. Phys. **54**, 5139 (1983).
[14] L. V. Keldysh, Zh. Eksp. Teor. Fiz. **48**, 1962 (1965) [Sov. Phys. JETP **21**, 1135 (1965)].
[15] Y. Wang and K. F. Brennan, J. Appl. Phys. **75**, 313 (1994).
[16] Y. Wang and K. F. Brennan, J. Appl. Phys. **76**, 974 (1994).
[17] N. Sano and A. Yoshii, Phys. Rev. B **45**, 4171 (1992).
[18] T. Kunikiyo, M. Takenaka, Y. Kamakura, M. Yamaji, H. Mizuno, M. Morifuji, K. Taniguchi, and C. Hamaguchi, J. Appl. Phys. **75**, 297 (1994).
[19] Y. Kamakura, H. Mizuno, M. Yamaji, M. Morifuji, K. Taniguchi, C. Hamaguchi, T. Kunikiyo, and M. Takenaka, J. Appl. Phys. **75**, 3500 (1994).
[20] M. Stobbe, A. Konies, R. Redmer, J. Henk, and W. Schattke, Phys. Rev. B **44**, 11 105 (1991).
[21] M. Stobbe, R. Redmer, and W. Schattke, Phys. Rev. B **49**, 4494 (1994).
[22] W. Quade, E. Scholl, and M. Rudan, Solid-State Electron. **36**, 1493 (1993).
[23] Y. C. Chang, D. Z.-Y. Ting, J. Y. Tang, and K. Hess, Appl. Phys. Lett. **42**, 76 (1983).
[24] W. Fawcett, A. D. Boardman, and S. Swain, J. Phys. Chem. Solids **31**, 1963 (1970).
[25] J. Y. Tang, Ph.D. thesis, University of Illinois, Urbana, IL, 1982.
[26] Y. Ohno, J. Appl. Phys. **64**, 4549 (1988).
[27] C. Jacoboni, C. Canali, G. Ottaviani, and A. A. Quaranta, Solid-State Electron. **20**, 77 (1977).
[28] K. F. Brennan, D. H. Park, K. Hess, and M. A. Littlejohn, J. Appl. Phys. **63**, 5004 (1988).
[29] K. F. Brennan, N. Mansour, and Y. Wang, Comput. Phys. Commun. **67**, 73 (1991).
[30] D. H. Park, Ph.D. thesis, Georgia Institute of Technology, Atlanta, GA, 1989.
[31] Landbolt-Börnstein, *Numerical Data and Functional Relationships in Science and Technology*, edited by K.-H. Hellwege and O. Madelung (Springer, New York, 1982).
[32] M. L. Cohen and T. K. Bergstresser, Phys. Rev. **141**, 789 (1966).
[33] R. van Overstraeten and H. DeMan, Solid-State Electron. **13**, 583 (1970).
[34] W. N. Grant, Solid-State Electron. **16**, 1189 (1973).
[35] M. H. Woods, W. C. Johnson, and M. A. Lambert, Solid-State Electron. **16**, 381 (1973).
[36] G. E. Bulman, V. M. Robbins, K. F. Brennan, K. Hess, and G. E. Stillman, IEEE Electron Device Lett. **EDL-4**, 181 (1983).
[37] N. Sano, T. Aoki, and A. Yoshii, Appl. Phys. Lett. **55**, 1418 (1989).
[38] N. Sano, M. Tomizawa, and A. Yoshii, Appl. Phys. Lett. **56**, 653 (1990).
[39] J. Bude and K. Hess, J. Appl. Phys. **72**, 3554 (1992).
[40] E. O. Kane, Phys. Rev. **159**, 624 (1967).
[41] E. Cartier, M. V. Fischetti, E. A. Eklund, and F. R. McFeely, Appl. Phys. Lett. **62**, 3339 (1993).
[42] Y. Wang and K. F. Brennan, J. Appl. Phys. **71**, 2736 (1992).
[43] M. G. Burt and C. Smith, J. Phys. C **17**, L47 (1984).
[44] M. G. Burt, S. Brand, C. Smith, and R. A. Abram, J. Phys. C **17**, 6385 (1985).
[45] J. Huldt, Phys. Status Solidi A **8**, 173 (1971).
[46] D. Hill and P. T. Landesberg, Proc. R. Soc. London Ser. A **347**, 547 (1976).
[47] D. Hill, Proc. R. Soc. London, Ser. A **347**, 565 (1976).
[48] J. Huldt, in 1972 International Conference on Semiconductor Physics, Warsaw, 1972, p. 1097.